

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 09:23:46 ; Search time 53.52 seconds
(without alignments)
566.576 Million cell updates/sec

Title: US-09-673-763-14
Perfect score: 1333
Sequence: 1 MTTPTLIIVPPSPAPSYSA.....QIRETLSPKSPSTKSS 273

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1333	100.0	273	20 AAY32176	Chlamydia trachoma
2	840	63.0	195	20 AAY37087	Amino acid sequenc
3	317	23.8	81	20 AAY37086	Amino acid sequenc
4	162.5	12.2	355	20 AAY32173	Chlamydia psittaci
5	154.5	11.6	397	20 AAY34783	Chlamydia pneumonia
6	124.5	9.3	2482	16 AAY72826	Human mitosisin. Ho
7	124.5	9.3	2482	19 AAY72826	Human mitosisin. Ho
8	124.5	9.3	3248	17 AAR99795	Kinetochore protei
9	120	9.0	1931	22 ABB61012	Drosophila melanog
10	116	8.7	875	22 AAE02245	Domestic mite Btll
11	116	8.7	878	22 AAE02242	Domestic mite Btll

12	116	8.7	1017	22 AAE02246	Domestic mite Btll
13	116	8.7	1177	22 AAB96721	Putative p. abysal
14	115.5	8.7	1372	19 AAW56473	Protein with Rho p
15	115	8.6	1325	18 AAW19540	Male-enhanced anti
16	115	8.6	1325	20 AAW94391	Mouse male enhance
17	113.5	8.5	881	22 ABG05280	Novel human diagno
18	113.5	8.5	881	22 ABG02058	Novel human diagno
19	112.5	8.4	622	22 ABB62816	Drosophila melanog
20	112	8.4	789	22 AAB95460	Human protein sequ
21	111.5	8.4	672	21 AAB21233	Corn MFPI. Zea ma
22	111.5	8.4	1192	22 AAU35310	Enterococcus faeca
23	111.5	8.4	1388	19 AAW56475	Protein with Rho p
24	111	8.3	334	15 AAR51227	Membrane antigen p
25	111	8.3	433	19 AAY20975	Human glial fibril
26	109.5	8.2	398	21 AAB21228	Partial protein en
27	109	8.2	2415	22 ABG20279	Novel human diagno
28	108.5	8.1	513	22 ABG08133	Novel human diagno
29	108.5	8.1	870	22 ABB70846	Drosophila melanog
30	108.5	8.1	1081	21 AAG50490	Arabidopsis thalia
31	108.5	8.1	1197	21 AAG50489	Arabidopsis thalia
32	108.5	8.1	1690	22 ABB61144	Drosophila melanog
33	108.5	8.1	1690	22 ABB61173	Drosophila melanog
34	108.5	8.1	2013	22 ABB62322	Drosophila melanog
35	108	8.1	430	21 AAG26548	Arabidopsis thalia
36	108	8.1	458	21 AAG26547	Arabidopsis thalia
37	108	8.1	473	21 AAG26546	Arabidopsis thalia
38	108	8.1	776	22 ABG05279	Novel human diagno
39	108	8.1	777	22 ABG20257	Novel human diagno
40	107.5	8.1	433	21 AAB58755	Breast and ovarian
41	107.5	8.1	469	18 AAW23820	Human sarcolectin.
42	107.5	8.1	469	21 AAY69289	Amino acid sequenc
43	107.5	8.1	546	21 AAY52397	Human keratin KERT
44	107.5	8.1	606	17 AAR99673	Receptor for hyalu
45	107.5	8.1	631	17 AAR99675	RHAMM 1-2a isoform

ALIGNMENTS

RESULT 1
AAY32176
ID AAY32176 standard; Protein: 273 AA.
XX
AC AAY32176;
XX
XX 01-FEB-2000 (first entry)
XX
DE Chlamydia trachomatis infection-specific protein Inca.
XX
KW Inca; infection; vaccine; therapy; diagnosis.
XX
OS Chlamydia trachomatis.
XX
PN WO9953948-A1.
XX
PD 28-OCT-1999.
XX
PF 20-APR-1999; 99WO-US08744.
XX
PR 20-APR-1998; 98US-0082438.
PR 21-APR-1998; 98US-0082588.
PR 22-MAY-1998; 98US-0086450.
XX
XX (UYOR-) UNIV OREGON STATE.
XX
XX Rocky DD, Bannantine JP;
XX
XX WPI; 1999-633904/54.
XX
XX N-PSDB; AAZ34590.
XX
XX Novel bacterial infection specific proteins for treating and diagnosing
XX
XX chlamydial infections

PS Claim 5; Page 45-46; 56pp; English.
XX This sequence represents novel infection-specific protein Inca of
CC Chlamydia trachomatis LGV-434 (serotype L2). Inca is found in the
CC inclusion membrane of infected cells. It is associated primarily
CC with the vegetative reticulate body form of Chlamydia rather than
CC with the refractile elementary body form. The invention includes: a
CC vaccine directed against the reticulate body form of Chlamydia
CC comprising 1 or more infection-specific proteins (see AAY32170-78),
CC including Inca, IncB and IncC; methods of using and producing such
CC a vaccine; methods for detection of infection-specific antibodies
CC or antigens in a biological specimen; and a method of using
CC therapeutic agents specifically directed against infection-specific
CC peptides, or the genes that code for such peptides, to treat
CC chlamydial infection.
XX
SQ Sequence 273 AA;

Query Match 100.0%; Score 1333; DB 20; Length 273;
Best Local Similarity 100.0%; Pred. No. 2.6e-106; Mismatches 0; Gaps 0;
Matches 273; Conservative 0; Indels 0;

QY 1 MTPTLIVIPPPAPPSYANRVPQPSLMDKIKKIAIASLILIGTIGTFLALGLHVLGFL 60
Db 1 mtptliivppspapsyanrvppslmdkikkiaiasliligtigflalghlvgl 60

QY 61 IAPQITIVLLALFITSAGNALYLOKTANLHLYQDLQREVGSKEINFMLSVLOKEFLHL 120
Db 61 iapqitivilalfitsagnalyloktanlhlyqdlqrevgskeinfmlsvlqkeflhl 120

QY 121 SKFATTSKDLASVQDFYSCLOGFRDNYKGFESLLDEYKNSTEEMRKLFQSEIIADLKG 180
Db 121 skfattskdlasvqdfysclogfrdnykgfeslldeyknsteemrklfesgeiialkg 180

QY 181 SVASLREIRFLTPLAEVRRRLAHNQESLTAATEELKTIRDSLRDEIGQLSKTLTTSQ 240
Db 181 svaslreirfltplaevrrrlahnqesltaateelktirdslrdeigqsktlttsq 240

QY 241 IALQKESSDLCQOIRETFLSSPRKSASPSTKSS 273
Db 241 ialqkessdlsqoiretflssprksaspstks 273

RESULT 2
AAY37087
AAY37087 standard; Protein; 195 AA.

AAY37087;
07-OCT-1999 (first entry)
Amino acid sequence of a Chlamydia trachomatis protein.
Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
paratrachoma; inclusion conjunctivitis; genital disease; perinephritis;
nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
bartholinitis; pneumopathy; venereal lymphogranulomatosis.
Chlamydia trachomatis.
WO9928475-A2.
10-JUN-1999.
27-NOV-1998; 98WO-IB01939.
04-NOV-1998; 98US-0107077.
28-NOV-1997; 97FR-0015041.
17-DEC-1997; 97FR-0016034.
(GEST) GENSET.

PI Griffais R;
XX
DR WPI; 1999-371125/31.
XX Genome sequence of Chlamydia trachomatis
XX
XX Disclosure; Page 892-893; 1755pp; English.
XX
CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as
CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
CC can also be used to control growth of the microorganism. Chlamydia
CC trachomatis is responsible for a large number of diseases, e.g. eye
CC diseases such as conventional trachoma, nonendemic trachoma,
CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
CC perinephritis, bartholinitis; pneumopathy in breast feeding infants;
CC and venereal lymphogranulomatosis. The polypeptides of the invention
CC may be of use in treating these diseases.
XX
SQ Sequence 195 AA;

Query Match 63.0%; Score 840; DB 20; Length 195;
Best Local Similarity 91.5%; Pred. No. 3.2e-64;
Matches 173; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY 64 QITIVLLALFITSAGNALYLOKTANLHLYQDLQREVGSKEINFMLSVLOKEFLHLSKE 123
Db 4 kftnvllaifitsagnalyloktanlhlyqdlqrevgskeinfmlsvlqkeflhlske 63

QY 124 FATTSKDLSAVSQDFYSCLOGFRDNYKGFESLLDEYKNSTEEMRKLFQSEIIADLKGSA 183
Db 64 fattskdlsavsqdfysclogfrdnykgfeslldeyknsteemrklfesgeiialkgsva 123

QY 184 SLREERFLTPLAEVRRRLAHNQESLTAATEELKTIRDSLRDEIGQLSKTLTTSOIAL 243
Db 124 slreerfltplaevrrrlahnqesltaateelktirdslrdeigqlsktlttsaklh 183

QY 244 QRKESDLC 252
Db 184 ynekragic 192

RESULT 3
AAY37086
ID AAY37086 standard; Protein; 81 AA.
XX
AC AAY37086;
XX
DT 07-OCT-1999 (first entry)
XX
DE Amino acid sequence of a Chlamydia trachomatis protein.
XX
KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW paratrachoma; inclusion conjunctivitis; genital disease; perinephritis;
KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.
XX
OS Chlamydia trachomatis.
XX
PN WO9928475-A2.
XX
PD 10-JUN-1999.
XX
PF 27-NOV-1998; 98WO-IB01939.
XX
PR 04-NOV-1998; 98US-0107077.
XX
PR 28-NOV-1997; 97FR-0015041.
XX
PR 17-DEC-1997; 97FR-0016034.
XX
PA (GEST) GENSET.
XX

PI Griffais R;
 XX WPI; 1999-371125/31.
 XX Genome sequence of Chlamydia trachomatis
 XX Disclosure; Page 892; 1755pp; English.
 XX AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
 CC of Chlamydia trachomatis (see AAZ01425). The polypeptides can be used as
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
 CC can also be used to control growth of the microorganism. Chlamydia
 CC trachomatis is responsible for a large number of diseases, e.g. eye
 CC diseases such as conventional trachoma, nonendemic trachoma, genital
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
 CC perinephritis, Bartholinitis; pneumopathy in breast feeding infants;
 CC and venereal lymphogranulomatosis. The polypeptides of the invention
 CC may be of use in treating these diseases.
 XX Sequence 81 AA;
 SQ
 Query Match 23.8%; Score 317; DB 20; Length 81;
 Best Local Similarity 98.5%; Pred. No. 7.2e-20;
 Matches 64; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MTTPTLIIVPPSPAPSANRVPOPSLMDKIKKIAATASLILIGTIGFLALLGHLVGF 60
 DB 16 mttptliivppspapsanrvpopsmdkikkiaaaslllgtigfllglnlvgl 75
 QY 61 IAPQI 65
 DB 76 iapqi 80
 RESULT 4
 AAY32173
 ID AAY32173 standard; Protein; 355 AA.
 XX
 AC AAY32173;
 XX
 DT 01-FEB-2000 (first entry)
 DE Chlamydia psittaci infection-specific protein Inca.
 KW Inca; infection; vaccine; therapy; diagnosis.
 XX Chlamydia psittaci.
 OS
 3 WO9953948-A1.
 PD 28-OCT-1999.
 XX
 PF 20-APR-1999; 99WO-US08744.
 XX
 PR 20-APR-1998; 98US-0082438.
 PR 21-APR-1998; 98US-0082588.
 PR 22-MAY-1998; 98US-0086450.
 XX
 PA (UYOR-) UNIV OREGON STATE.
 XX
 XX Rocky DD, Bannantine JP;
 XX WPI; 1999-633904/54.
 DR N-PSDB; AAZ34587.
 XX
 XX Novel bacterial infection specific proteins for treating and diagnosing
 PT chlamydial infections
 XX
 PS Claim 5; Page 39-40; 56pp; English.
 XX
 CC This sequence represents novel infection-specific protein Inca of

CC Chlamydia psittaci strain GPIC. Inca is found in the inclusion
 CC membrane of infected cells. It is associated primarily with the
 CC vegetative reticulate body form of Chlamydia rather than with the
 CC refractile elementary body form. The invention includes: a vaccine
 CC directed against the reticulate body form of Chlamydia comprising 1
 CC or more infection-specific proteins (see AAY32170-78), including
 CC Inca, IncB and IncC; methods of using and producing such a vaccine;
 CC methods for detection of infection-specific antibodies or antigens
 CC in a biological specimen; and a method of using therapeutic agents
 CC specifically directed against infection-specific peptides, or the
 CC genes that code for such peptides, to treat chlamydial infection.
 XX Sequence 355 AA;
 SQ
 Query Match 12.2%; Score 162.5; DB 20; Length 355;
 Best Local Similarity 21.4%; Pred. No. 9.3e-06;
 Matches 63; Conservative 61; Mismatches 111; Indels 59; Gaps 8;
 QY 8 VIPSPAPSPAPSANRV-----POPSLMDKIKKIAATASLILIGTIGFLALLGHLVG 58
 DB 32 lipisieaptssaaavgaaktaiepgsrpllqricyivkkaiaaiaifvvgaaalvclylg 91
 QY 59 FLIAPQITIVLLALFITSLSA-----GNALYLQKTANLHLXYQDLQR-----EV 100
 DB 92 svistpsliimlaamlvsvivitaigrdtpsgvvr-----hmkqqiqqfgeentrlhtav 147
 QY 101 GSKKEINFMLS--VLOKEFLH-----LSKEFATSKDLSAVSQDFYSCLOGFRDNYKGFE 153
 DB 148 enlkavnvelseqinqlkqhtrlsdfgrleantgofatlaadfqisleeefksvgtkve 207
 QY 154 SLIDYEKNSTEEMKLFQSEIIADLKGVSASLREIEIRFLPLAEVRRRLAHNQESLTAI 213
 DB 208 tmlspfeklaqsketfsgavqammsvvtelrt-----nlnakell 250
 QY 214 EELKTIIRDSLRDEIGQLSQLSKLTLSQIALQKESDLCQIRETLSSPRKAS 267
 DB 251 tenktvleqlkad----agireeqvrflekkrkdeleacstlshsiatiqestt 300
 RESULT 5
 AAY34783
 ID AAY34783 standard; Protein; 397 AA.
 XX
 AC AAY34783;
 XX
 DT 13-SEP-1999 (first entry)
 DE Chlamydia pneumoniae transmembrane protein sequence.
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope.
 XX Chlamydia pneumoniae.
 OS
 XX WO9927105-A2.
 XX
 PD 03-JUN-1999.
 XX
 PF 20-NOV-1998; 98WO-IB01890.
 XX
 PR 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 XX
 XX (GEST) GENSET.
 PA
 XX Griffais R;
 XX WPI; 1999-357842/30.
 XX
 XX Genome sequence of Chlamydia pneumoniae
 XX

FH Key Location/Qualifiers
 FT 258..280
 FT /note= "leucine heptad repeat"
 FT 340..362
 FT 564..593
 FT 1387..1443
 FT 1885..1962
 FT 2146..2188
 FT 2165..2187
 FT /note= "leucine heptad repeat"
 FT Misc-difference 2188
 FT Misc-difference 2300
 FT /label= "Bipartite targeting motif"
 FT /note= "Optionally C or G"
 FT Misc-difference 2189
 FT Misc-difference 2301
 FT Misc-difference 2303
 FT /label= "Bipartite targeting motif"
 FT /note= "Optionally A or T"
 XX US5710022-A.
 XX 20-JAN-1998.
 XX 24-OCT-1994; 94US-0328254.
 XX 24-OCT-1994; 94US-0328254.
 XX 22-OCT-1993; 93US-0141239.
 XX (TEXA) UNIV TEXAS SYSTEM.
 XX Lee W, Zhu X;
 XX WPI; 1998-109817/10.
 XX N-PSDB; AAV09076.
 XX New isolated mitosis protein and gene - useful for, e.g. developing
 XX products for therapy and diagnosis of hyper-proliferative disorders
 XX such as cancers or psoriasis
 XX Claim 1; Column 40-52; 43pp; English.
 XX This is the amino acid sequence for mitosis, a phosphoprotein
 XX necessary for the cell to enter mitosis. The protein's degradation is
 XX also necessary for the cell to advance into the next stages of mitosis.
 XX The mitosis protein, can be used to control the growth of cells. An
 XX anti-mitosis antibody, a mutant or a non-functional analogue of mitosis
 XX can inhibit the mitotic cell cycle by preventing the cells from entering
 XX the M phase, and over expression of mitosis or its functional
 XX equivalent, would inhibit the cycle by preventing cells from leaving the
 XX M phase. Antagonists to this protein can be used to control
 XX hyperproliferative cells in, (e.g. thyroid hyperplasia, Grave's disease,
 XX psoriasis, benign prostatic hypertrophy, Li-Fraumeni syndrome, breast
 XX cancer, sarcomas and other neoplasms, bladder cancer, colon cancer,
 XX lung cancer and various leukemias and lymphomas). Reintroduction or
 XX supplementation of lost mitosis function by introduction of the protein
 XX or nucleic acid encoding the protein into a cell can restore defective
 XX chromosome segregation, which is a marker of progressing malignancy.
 XX Malignant proliferation of cells can then be halted. The protein
 XX can also be used for the detection and diagnosis of hyperproliferative
 XX cells.
 XX Sequence 2482 AA;

Query Match 9.3%; Score 124.5; DB 19; Length 2482;
 Best Local Similarity 26.2%; Pred. No. 0.23;
 Matches 56; Conservative 32; Mismatches 83; Indels 43; Gaps 8;

OY 77 LAGNALYLQK-----TANLHLVQDQLQREVGLSKLEINFMLSVLQKFLHLSKEFAT--- 126
 DB 314 llsletislekemasiislnkreieeltqngtltkeinaslnqemlnliqsfsfanyid 373

OY 127 -TSKDL SAVS-----QDFYSCLOGFRDNYKGFESLLDEYKNSTEEMRKL-----FSOEIIA 176
 DB 374 ereksiselsdqykqeklllllqrceetngayedlsqkykaaqeknskllnectslce 433
 OY 177 DLKGSVASLREEI-----RFLPLAEVVRRLAHNQESLTAATIEELKTIRDSLRDEI----- 227
 DB 434 nrkneleqkkaefakehgefllkfaeer---ngnlml-----eietvqgairstmdnq 486
 OY 228 -----GQLSKLTLSQIALQRKESDLCQ 254
 DB 487 nnskseagglkqeimtlkeeqnkmqkevndllqe 520
 RESULT 8
 AAR99795
 ID AAR99795 standard; Protein; 3248 AA.
 XX
 AC AAR99795;
 XX
 XX 08-OCT-1996 (first entry)
 XX Kinetochores protein CENP-F.
 XX Kinetochores protein; CENP-F; cell cycle; cancer; diagnosis;
 XX autoimmune antibody.
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT Domain 1..200
 FT /label= Extended_coiled_structure
 FT 280..1350
 FT /label= Extended_coiled_structure
 FT 1380..1610
 FT /label= Globular_domain
 FT /note= "globular domain consists of 2 direct
 FT repeats of 95 amino acids"
 FT Domain 1620..1750
 FT /label= Extended_coiled_structure
 FT 1850..2990
 FT /label= Extended_coiled_structure
 FT 3048..3248
 FT /label= C-terminal_domain
 FT /note= "the C-terminal domain is predicted to
 FT form a proline-rich (10.8%) highly
 FT basic (pI 10) globular domain"
 XX
 PN W09617867-A1.
 XX 13-JUN-1996.
 XX 08-DEC-1995; 95WO-US16216.
 XX 09-DEC-1994; 94US-0353700.
 XX (FOXC-) FOX CHASE CANCER CENT.
 XX (UVTE-) UNIV TECHNOLOGIES INT INC.
 XX Rattner JB, Yen TJ;
 XX WPI; 1996-287116/29.
 XX N-PSDB; AAT34578.
 XX DNA encoding kinetochores protein - used as a marker for the G2 and M
 XX phases of a cell cycle, partic. for detection of malignant diseases
 PS Claim 12; Page 41-54; 72pp; English.
 CC A 372 kDa human kinetochores protein, CENP-F (AAR99795), is detected
 CC by immunofluorescence microscopy only during the G2 and M phases
 CC of a cell cycle. It is the product of a cDNA clone (AAR34578)
 CC isolated from a breast carcinoma cDNA library. Recombinant CENP-F
 CC can be produced by expression in prokaryotic or eukaryotic host

CC Immunogenic protein is useful for preventing, reducing or ameliorating
 CC Blomia tropicalis hypersensitivity condition such as atopic dermatitis,
 CC immediate hypersensitivity, systemic anaphylaxis, allergic rhinitis or
 CC asthma and for modulating an immune response directed to Bt allergen in
 CC a subject. The Bt allergens are also useful for detecting antibody
 CC directed to all or a part of Bt allergen in a biological sample from a
 CC subject. Antibodies to Bt allergens are also used as therapeutic or
 CC diagnostic agents, to screen Bt immunoassays and as antagonists to
 CC inhibit Bt activity under circumstances where temporary hypersensitivity
 CC inhibition is required. The present sequence is Bt11 allergen.
 XX
 SQ Sequence 875 AA;

Query Match 8.7%; Score 116; DB 22; Length 875;
 Best Local Similarity 26.9%; Pred. No. 0.31;
 Matches 46; Conservative 31; Mismatches 72; Indels 22; Gaps 5;
 QY 94 QDLOREVGSLKEINFMFLSVLQKEFLHLSKEFATTSKDLSAVSQDFYSCLOGFRDNYKGE 153
 Db 364 qalekrvsqlekinidkkskleevsmleq---tqkdlrvkiad---lqklqhey---e 413
 154 SLLDEYKNSSTEMRKLFQOEIIADLKGSVASLREIEIRFLTPLAEEVRLAHNQESLTAAI 213
 Db 414 kirdqkealarenkklad-----dlaeaqsqndahrriheqeieikrlenereslaaay 468
 QY 214 BELKTIKRLDRDEIGQLSQTSLTSLQALQKESDLCQIRETLSSPRK 264
 Db 469 keaetlirk-----qeeaknqrtaelaqtrhdyekrlaqkeeeiealrk 512

RESULT 11
 AAE02242
 ID AAE02242 standard; Protein; 878 AA.
 AC AAE02242;
 DT 31-JUL-2001 (first entry)
 DE Domestic mite Bt11 allergen #7.
 KW Mite; immunogenic protein; Bt allergen; therapy; atopic dermatitis;
 KW immediate hypersensitivity; systemic anaphylaxis; allergic rhinitis;
 KW asthma; anti-allergic; anti-inflammatory; immunosuppressive.
 OS Blomia tropicalis.
 XX WO200130817-A1.
 XX 03-MAY-2001.

10-OCT-2000; 2000WO-AU01227.
 XX 26-OCT-1999; 99SG-0005313.
 PR 18-JUL-2000; 2000AU-0008842.
 PR 18-JUL-2000; 2000AU-0008844.
 PR 18-JUL-2000; 2000AU-0008845.
 XX (UYSI-) UNIV SINGAPORE NAT.
 PA Chua KY, Cheong N, Lee BW;
 XX WPI: 2001-308609/32.
 DR N-PSDB; AAD06236.
 XX

Novel immunogenic protein derived from house mite, Blomia tropicalis
 useful for treating and diagnosing conditions involving induction of
 PT immuneresponse to mite, such as allergic asthma, atopic dermatitis,
 PT rhinitis
 XX

Claim 4; Fig 3; 230pp; English.

The present invention relates to immunogenic proteins, referred as Bt

CC allergen, is derived from domestic mite, Blomia tropicalis. The specific
 CC Bt allergens of the invention includes Bt11, Bt10, Bt5 and BtA2. The
 CC immunogenic protein is useful for preventing, reducing or ameliorating
 CC Blomia tropicalis hypersensitivity condition such as atopic dermatitis,
 CC immediate hypersensitivity, systemic anaphylaxis, allergic rhinitis or
 CC asthma and for modulating an immune response directed to Bt allergen in
 CC a subject. The Bt allergens are also useful for detecting antibody
 CC directed to all or a part of Bt allergen in a biological sample from a
 CC subject. Antibodies to Bt allergens are also used as therapeutic or
 CC diagnostic agents, to screen Bt immunoassays and as antagonists to
 CC inhibit Bt activity under circumstances where temporary hypersensitivity
 CC inhibition is required. The present sequence is Bt11 allergen.
 XX
 SQ Sequence 878 AA;

Query Match 8.7%; Score 116; DB 22; Length 878;
 Best Local Similarity 26.9%; Pred. No. 0.31;
 Matches 46; Conservative 31; Mismatches 72; Indels 22; Gaps 5;
 QY 94 QDLOREVGSLKEINFMFLSVLQKEFLHLSKEFATTSKDLSAVSQDFYSCLOGFRDNYKGE 153
 Db 367 qalekrvsqlekinidkkskleevsmleq---tqkdlrvkiad---lqklqhey---e 416
 154 SLLDEYKNSSTEMRKLFQOEIIADLKGSVASLREIEIRFLTPLAEEVRLAHNQESLTAAI 213
 Db 417 kirdqkealarenkklad-----dlaeaqsqndahrriheqeieikrlenereslaaay 471
 QY 214 BELKTIKRLDRDEIGQLSQTSLTSLQALQKESDLCQIRETLSSPRK 264
 Db 472 keaetlirk-----qeeaknqrtaelaqtrhdyekrlaqkeeeiealrk 515

RESULT 12
 AAE02246
 ID AAE02246 standard; Protein; 1017 AA.
 AC AAE02246;
 DT 31-JUL-2001 (first entry)
 DE Domestic mite Bt11 allergen polymorphic variant.
 KW Mite; immunogenic protein; Bt allergen; therapy; atopic dermatitis;
 KW immediate hypersensitivity; systemic anaphylaxis; allergic rhinitis;
 KW asthma; anti-allergic; anti-inflammatory; immunosuppressive.
 OS Blomia tropicalis.

Key Location/Qualifiers
 FT Misc-difference 41 /note= "Encoded by TAG"
 FT Misc-difference 42 /note= "Encoded by TAG"
 FT Misc-difference 56 /note= "Encoded by TGA"
 FT Misc-difference 71 /note= "Encoded by TAA"
 FT Misc-difference 76 /note= "Encoded by TAG"
 FT Misc-difference 80 /note= "Encoded by TGA"
 FT Misc-difference 86 /note= "Encoded by TAA"
 FT Misc-difference 965 /note= "Encoded by TAA"
 FT Misc-difference 998 /note= "Encoded by TAA"
 XX WO200130817-A1.
 XX 03-MAY-2001.

XX PF 10-OCT-2000; 2000WO-AU01227.
XX PR 26-OCT-1999; 99SG-0005313.
XX PR 18-JUL-2000; 2000AU-0008842.
XX PR 18-JUL-2000; 2000AU-0008844.
XX PR 18-JUL-2000; 2000AU-0008845.
XX (UYSI-) UNIV SINGAPORE NAT.
XX PA Chua KY, Cheong N, Lee BW;
XX PI WPI; 2001-308609/32.
XX DR N-PSDB; AAD06245.
XX Novel immunogenic protein derived from house mite, Blomia tropicalis
XX useful for treating and diagnosing conditions involving induction of
XX immuneresponse to mite, such as allergic asthma, atopic dermatitis,
XX rhinitis -
XX Claim 6; Fig 7; 230pp; English.
XX The present invention relates to immunogenic proteins, referred as Bt
XX allergen, is derived from domestic mite, Blomia tropicalis. The specific
XX Bt allergens of the invention includes Bt11, Bt10, Bt5 and BtA2. The
XX immunogenic protein is useful for preventing, reducing or ameliorating
XX Blomia tropicalis hypersensitivity condition such as atopic dermatitis,
XX immediate hypersensitivity, systemic anaphylaxis, allergic rhinitis or
XX asthma and for modulating an immune response directed to Bt allergen in
XX a subject. The Bt allergens are also useful for detecting antibody
XX directed to all or a part of Bt allergen in a biological sample from a
XX subject. Antibodies to Bt allergens are also used as therapeutic or
XX diagnostic agents, to screen Bt immunoassays and as antagonists to
XX inhibit Bt activity under circumstances where temporary hypersensitivity
XX inhibition is required. The present sequence is a protein encoded
XX by Bt11 polymorphic variant.
XX SQ Sequence 1017 AA;
Query Match 8.7%; Score 116; DB 22; Length 1017;
Best Local Similarity 26.9%; Pred. No. 0.37;
Matches 46; Conservative 31; Mismatches 72; Indels 22; Gaps 5;
QY 94 QDLQREVGSLKEINFMLSVLOKEFLHLSKEEFATTSKDSAVSQDFYSCLOQFRDNYKGF 153
Db 453 qalekvtvsglekinidkksleevsmllqeq---tqkdlrvkiad---lqklqhey---e 502
154 SLDEYKNSTEMRKLFQSEIADLKGVSALREERFRLTPLAEVRRRLAHNQESLTAI 213
Db 503 klrdqkealarenkklad-----dlaeaqsqindahrrihegeleikrlenereelaay 557
QY 214 EELKTIKRLSRDEIGQLSOLSKTLTSLQIALORKESDLSQIRETLSSPRK 264
Db 558 keaatlirk-----qeeaknqlrtaelaqttrhdyekrlaqkeeelealrk 601
RESULT 13
AAB96721
ID AAB96721 standard; Protein; 1177 AA.
XX AC AAB96721;
XX DT 29-OCT-2001 (first entry)
XX DE Putative P. abyssi ATPase involved in DNA repair #3.
XX Hyperthermophilic archaeon; hyperthermophilic protein.
XX KW Pyrococcus abyssi.
XX OS FR2792651-A1.
XX PN 27-OCT-2000.
XX PD
XX PF 21-APR-1999; 99FR-0005034.
XX PR 21-APR-1999; 99FR-0005034.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX (IFRE-) IFREMER INST FR RECH EXPL MER.
XX Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
XX Querellou J, Weissenbach J, Saurin W, Heilig R;
XX WPI; 2001-126236/14.
XX New nucleotide sequences isolated from Pyrococcus abyssi encode
XX proteins useful in industry -
XX Claim 7; Pages 1483-1487; 1657pp; French.
XX The present invention relates to the genomic sequence of Pyrococcus
XX abyssi (see AAF86431 and AAF41223-7) and P. abyssi proteins. P. abyssi is
XX a hyperthermophilic archaeon, which is isolated from deep-sea
XX hydrothermal vents. The present sequence is one such P. abyssi protein.
XX The proteins of the present invention have various potential industrial
XX uses, since the proteins are stable at very high temperatures, some up to
XX 110 degrees centigrade.
XX Note: This patent is in the same patent family as WO2000065062, which
XX contains additional sequences as shown in AAB99132-AAB99143,
XX AAB75903-AAH75920 and AAG66436.
XX SQ Sequence 1177 AA;
Query Match 8.7%; Score 116; DB 22; Length 1177;
Best Local Similarity 19.7%; Pred. No. 0.45; 72; Indels 82; Gaps 9;
Matches 48; Conservative 42; Mismatches 72; Indels 82; Gaps 9;
QY 82 LYLOKANTLHLYQDLQREVGSLK-----EINFMLSVLOKEFLHLSKEFATTSKD 130
Db 673 lklrkea-----leaeinslkvelgnggfelrikmseikeitlltrdieklse 725
QY 131 LSAVSQDFYSCLOQFRDNYKGFESL-----LDEYKNSTEMRKLFQSE- 172
Db 726 erliikse-----ledsqgleidriihekkgelakrlgrkierlerkrklkalenp 778
QY 173 -----ETIADLKGVSALREEI-----RFLTP-----LAEEVR----- 200
Db 779 earevtekirovegeigklreelsrvesrleslnsrlnneeliprkasleeeleglvnkin 838
QY 201 -----RLAHNQE---SLTAAIEELKTIKRLSRDEIGQLSOLSKTLTSLQIALORKESDLS 253
Db 839 alkaniaeneevlkgkgleelkakeesvhsksieyrrkrkeelekeirelkekeelsk 898
QY 254 QIRE 257
Db 899 rmqge 902
RESULT 14
AAW56473
ID AAW56473 standard; Protein; 1372 AA.
XX AC AAW56473;
XX DT 14-AUG-1998 (first entry)
XX DE Protein with Rho protein-combining and kinase activity.
XX Rho protein-binding activity; protein kinase activity; inhibitor;
XX smooth muscle fibre formation; smooth muscle contraction;
XX circulatory disease; treatment; tumour formation; metastasis inhibitor;
XX autoimmune disease; platelet aggregation inhibitor.
XX Bos sp.

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	124.5	9.3	2482	1	US-08-328-254-6		Sequence 6, Appli
2	124.5	9.3	3248	1	US-08-353-700-1		Sequence 1, Appli
3	124.5	9.3	3248	5	PCT-US95-16216-1		Sequence 1, Appli
4	117	8.8	775	1	US-07-603-133B-16		Sequence 16, Appli
5	115.5	8.7	1388	2	US-08-685-576-1		Sequence 1, Appli
6	113.5	8.5	775	1	US-07-603-133B-15		Sequence 15, Appli
7	111.5	8.4	1388	2	US-08-685-576-4		Sequence 4, Appli
8	107.5	8.1	546	2	US-09-067-351-1		Sequence 1, Appli
9	107.5	8.1	546	4	US-09-360-430-1		Sequence 1, Appli
10	106	8.0	477	1	US-08-402-217A-3		Sequence 1, Appli
11	106	8.0	477	1	US-08-700-178-3		Sequence 3, Appli
12	106	8.0	477	3	US-08-995-654-3		Sequence 3, Appli
13	106	8.0	604	2	US-08-468-576B-12		Sequence 3, Appli
14	106	8.0	604	2	US-08-468-579B-12		Sequence 12, Appli
15	106	8.0	604	3	US-08-468-577B-12		Sequence 12, Appli
16	106	8.0	2101	1	US-08-466-390-4		Sequence 12, Appli
17	106	8.0	2101	1	US-08-470-950-4		Sequence 4, Appli
18	106	8.0	2101	1	US-08-467-781-4		Sequence 4, Appli
19	106	8.0	2101	1	US-08-195-487-4		Sequence 4, Appli
20	106	8.0	2101	2	US-08-483-824-4		Sequence 4, Appli
21	106	8.0	2101	4	US-09-452-294-1		Sequence 4, Appli
22	106	8.0	2101	5	PCT-US93-06160-4		Sequence 4, Appli
23	104	7.8	1354	3	US-08-685-871-2		Sequence 1, Appli
24	103.5	7.8	712	2	US-08-468-576B-17		Sequence 2, Appli
25	103.5	7.8	712	3	US-08-468-579B-17		Sequence 17, Appli
26	103.5	7.8	712	3	US-08-468-577B-17		Sequence 17, Appli
27	103	7.7	747	3	US-08-089-397A-16		Sequence 16, Appli

Db 314 LLSETLSLEKKEWSSIIISLNKREIBELTQENGTLKEINASLNQKMNLIQKSEFANYID 373
Qy 127 -TSKDLASVS-----QDFYSCLOGFRDNYKGFESLLDDEYKNSTEEMRKL-----FSQEIIA 176
Db 374 EREKSISLSLDYQKOEKLLILQRCETGNAYEDLSQYKAAQEKNSKLECLLNECTSLCE 433
Qy 177 DLKGSVASLREEI-----RFLTPLAEVRRLAHQESLTAATEELKTRDSLRDEI----- 227
Db 434 NRKNELEQLKEAFKAEHQEFLTKLAFABER---NONML-----ELETVOQALRSEMTDQ 486
Qy 228 -----GOLSQSLKTLTSQIALQKESSDLCQ 254
Db 487 NNSKSEAGGLKQEIIMTLKEQNKMQKEVNDLLOE 520
RESULT 2
PCT-US95-700-1
Sequence 1, Application US/08353700
Patent No. 5599919
GENERAL INFORMATION:
APPLICANT: YEN, TIMOTHY J.
TITLE OF INVENTION: NUCLEIC ACID ENCODING A
TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
STREET: 1601 MARKET STREET, SUITE 720
CITY: PHILADELPHIA
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,700
FILING DATE: 09-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: REED, JANET E.
REGISTRATION NUMBER: 36,252
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE: HUMAN
ORGANISM: HUMAN
US-08-353-700-1
Query Match 9.3%; Score 124.5; DB 1; Length 3248;
Best Local Similarity 26.2%; Pred. No. 0.034;
Matches 56; Conservative 32; Mismatches 83; Indels 43; Gaps 8;
Qy 77 LAGNALYLOK-----TANLHYODLOREVGSLSKAEINFMLSVLQKFEFLHLSKEFAT--- 126
Db 946 LLSETLSLEKKEWSSIIISLNKREIBELTQENGTLKEINASLNQKMNLIQKSEFANYID 1005
Qy 127 -TSKDLASVS-----QDFYSCLOGFRDNYKGFESLLDDEYKNSTEEMRKL-----FSQEIIA 176
Db 1006 EREKSISLSLDYQKOEKLLILQRCETGNAYEDLSQYKAAQEKNSKLECLLNECTSLCE 1065

Qy 177 DLKGSVASLREEI-----RFLTPLAEVRRLAHQESLTAATEELKTRDSLRDEI----- 227
Db 1066 NRKNELEQLKEAFKAEHQEFLTKLAFABER---NONML-----ELETVOQALRSEMTDQ 1118
Qy 228 -----GOLSQSLKTLTSQIALQKESSDLCQ 254
Db 1119 NNSKSEAGGLKQEIIMTLKEQNKMQKEVNDLLOE 1152
RESULT 3
PCT-US95-16216-1
Sequence 1, Application PC/TUS9516216
GENERAL INFORMATION:
APPLICANT: Yen, Timothy J.
TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16216
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,700
FILING DATE: 09-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
PCT-US95-16216-1
Query Match 9.3%; Score 124.5; DB 5; Length 3248;
Best Local Similarity 26.2%; Pred. No. 0.034;
Matches 56; Conservative 32; Mismatches 83; Indels 43; Gaps 8;
Qy 77 LAGNALYLOK-----TANLHYODLOREVGSLSKAEINFMLSVLQKFEFLHLSKEFAT--- 126
Db 946 LLSETLSLEKKEWSSIIISLNKREIBELTQENGTLKEINASLNQKMNLIQKSEFANYID 1005
Qy 127 -TSKDLASVS-----QDFYSCLOGFRDNYKGFESLLDDEYKNSTEEMRKL-----FSQEIIA 176
Db 1006 EREKSISLSLDYQKOEKLLILQRCETGNAYEDLSQYKAAQEKNSKLECLLNECTSLCE 1065
Qy 177 DLKGSVASLREEI-----RFLTPLAEVRRLAHQESLTAATEELKTRDSLRDEI----- 227
Db 1066 NRKNELEQLKEAFKAEHQEFLTKLAFABER---NONML-----ELETVOQALRSEMTDQ 1118
Qy 228 -----GOLSQSLKTLTSQIALQKESSDLCQ 254

1

: Sequence 15 Application US/07603133B
 : Patent No. 5298244
 : GENERAL INFORMATION:

GENERAL INFORMATION:
 ; APPLICANT: Kaibuchi, Kozo
 ; APPLICANT: Iwamatsu, Akihiro

us-09-673-763-14.ra

Tue Aug 13 14:23:40 2002

```

;
; APPLICANT: Redmond, Mark J.
; APPLICANT: Ijaz, Mohammed K.
; APPLICANT: Parker, Michael D.
; TITLE OF INVENTION: ASSEMBLED VIRAL PARTICLES AND THEIR
; TITLE OF INVENTION: USE IN A VACCINE TO ROTAVIRAL DISEASE
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 545 Middlefield Road, Suite 200
; CITY: Menlo Park
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/603,133B
; FILING DATE: 19901025
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Robins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 9313-0004.00
; TELEPHONE: (415) 327-7250
; TELEFAX: (415) 327-2951
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 775 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
;
US-07-603-133B-15

```

```

Query Match      8.5%; Score 113.5; DB 1; Length 775;
Best Local Similarity 23.2%; Pred. No. 0.042;
Matches 55; Conservative 45; Mismatches 100; Indels 37; Gaps 9;

QY 43 LIGTIGLALLGVGLIAPQITIVLLALFITSAGNALYLQKTAN-LHLVQDLQREV 101
DB 446 LVLGPAFPNPSGHEY-YETAGRFSEFILL-----VPSNDDYQTPIMNSVTVRQDLEROLG 498
102 SLKEINFMLSVLQKFEFLHLSKEFATSK-DLSAVSQDFYSCLOGFRDNYKFESILLDEYK 160
DB 499 DLRE-----EFNSLSQEIAMTQILDALLPLDMFSM-----FSGIKSTIDAAK 541
QY 161 NTEEMRKLFQSE---ITADLKGVSASLREIRFLTPLAEVRRRLAHNQESLTAAIL 216
DB 542 SMATKVMKPKRSGLAGTSISELTGSLNNAASSIRSSSIRSNISSIS----VWTDVSEQI 597
QY 217 KTIKRLRDEIGQLSQTLSQTALQKES---SDLSQIRETSLSPRKSPASPT 270
DB 598 AGSSDSVNSISQMSAISRLRLREITTTQTEGMNFDISAAVLKTKIDRSTHISPT 654

RESULT 7
US-08-685-576-4
; Sequence 4, Application US/08685576
; Patent No. 5906819
; GENERAL INFORMATION:
; APPLICANT: Kaibuchi, Kojo
; APPLICANT: Iwamatsu, Akihiro
; APPLICANT: Nakano, Takeshi
; APPLICANT: Ito, Masaaki
; APPLICANT: Takahashi, No. 5906819uaki
; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
; NUMBER OF SEQUENCES: 16

```

```

;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,576
; FILING DATE: 24-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-325129
; FILING DATE: 20-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-17150
; FILING DATE: 05-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-131206
; FILING DATE: 26-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16887/843
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1388 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-685-576-4

```

```

Query Match      8.4%; Score 111.5; DB 2; Length 1388;
Best Local Similarity 22.1%; Pred. No. 0.15;
Matches 43; Conservative 40; Mismatches 91; Indels 21; Gaps 4;

QY 84 LQKTANLHLVQDLQREVGSLEK-EINEMLSVLQKFEFLHLSKEFATSKDLSAVSQDFYSCL 142
DB 667 LEKRLQLOERFTDLEKEKSNMEIDMTYQLKVIQOLESLEQEEAEHAKATKARLADKNKIYESIE 726
QY 143 QGFRDNYKGFESLLDEYKKNSTEEMRKLF-----SQEIIADLKGVSASLR 186
DB 727 EAKSEAMKEMKKLEERTLKQKVENLLLEAEKRCSLDDCKLQSQOKINELLKQKDVLN 786
QY 187 BEIRFLT-PLAEVRRRLAHNQESLTAAILKLTIRDS---LRDEIGQLSQTLSQTIA 242
DB 787 EDVRNLTLKIEQTKRCLTQNDLKMOTQVNTLKMSEKQLQENHLMEMKMLKONA 846
QY 243 LQKRESSDLCQIRE 257
DB 847 ELRKRQDQDGQMK 861

RESULT 8
US-09-067-351-1
; Sequence 1, Application US/09067351
; Patent No. 5994081
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Baughn, Mariah
; TITLE OF INVENTION: HUMAN KERATINS

```

NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/067,351
FILING DATE: Herewith
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: CERRONE, MICHAEL C.
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0511 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 546 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: PANCUT02
CLONE: 1467090
US-09-067-351-1

Query Match 8.1%; Score 107.5; DB 2; Length 546;
Best Local Similarity 24.1%; Pred. No. 0.091;
Matches 48; Conservative 31; Mismatches 73; Indels 47; Gaps 7;

QY 80 NALYLQKTANLHLYQDLQREVGLK-EINFMLSVLQKEFLHLSKEFATTSKDL SAVSQDF 138
Db 279 DAAYMSKV-----ELEAKVDALNDEINFLRLNETELQSQISDTSVVL----- 325
QY 139 YSCLOGFRDNYKG--FESLLDEYKNSTEMRKLFSOEIIA-----DL 178
Db 326 -----MNSRSLDLDGIIAEVKAQEYEMAKCSRAEAAYQTKFETLQAQAGKHGDDL 378
QY 179 KGSVASLREEIFLTPLAEEVRLAHNQESLTAATIEELKTIRDSLREDTGQLSKLTIT 238
379 RNRTRNEISEMNRAIQRLQAEIDNKNQRAKLEAAIAEAE-----RGELALKDARAKQEE 433
QY 239 SQIALQKRESSDLCSQIRE 257
Db 434 LEAALQRAK-QDMARQLRE 451

RESULT 9
US-09-360-490-1
Sequence 1, Application US/09360490
Patent No. 6221843
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Hillman, Jennifer L.
APPLICANT: Corley, Neil C.
APPLICANT: Baughn, Mariah
TITLE OF INVENTION: HUMAN KERATINS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA

COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/360,490
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/067,351
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: CERRONE, MICHAEL C.
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0511 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 546 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: PANCUT02
CLONE: 1467090
US-09-360-490-1

Query Match 8.1%; Score 107.5; DB 4; Length 546;
Best Local Similarity 24.1%; Pred. No. 0.091;
Matches 48; Conservative 31; Mismatches 73; Indels 47; Gaps 7;

QY 80 NALYLQKTANLHLYQDLQREVGLK-EINFMLSVLQKEFLHLSKEFATTSKDL SAVSQDF 138
Db 279 DAAYMSKV-----ELEAKVDALNDEINFLRLNETELQSQISDTSVVL----- 325
QY 139 YSCLOGFRDNYKG--FESLLDEYKNSTEMRKLFSOEIIA-----DL 178
Db 326 -----MNSRSLDLDGIIAEVKAQEYEMAKCSRAEAAYQTKFETLQAQAGKHGDDL 378
QY 179 KGSVASLREEIFLTPLAEEVRLAHNQESLTAATIEELKTIRDSLREDTGQLSKLTIT 238
379 RNRTRNEISEMNRAIQRLQAEIDNKNQRAKLEAAIAEAE-----RGELALKDARAKQEE 433
QY 239 SQIALQKRESSDLCSQIRE 257
Db 434 LEAALQRAK-QDMARQLRE 451

RESULT 10
US-08-402-217A-3
Sequence 3, Application US/08402217A
Patent No. 5587301
GENERAL INFORMATION:
APPLICANT: Hawkins, Phillip R.
APPLICANT: Wilde, Craig G.
APPLICANT: Seilhamer, Jeffrey J.
TITLE OF INVENTION: HYALURONAN RECEPTOR EXPRESSED IN HUMAN
TITLE OF INVENTION: UMBILICAL VEIN ENDOTHELIAL CELLS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/402,217A
FILING DATE: 10-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Luther, Barbara J.
REGISTRATION NUMBER: 33954
REFERENCE/DOCKET NUMBER: PF-002808
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-852-0195
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
IMMEDIATE SOURCE:
LIBRARY: mouse
CLONE: GI 53979
US-08-402-217A-3

Query Match 8.0%; Score 106; DB 1; Length 477;
Best Local Similarity 23.9%; Pred. No. 0.1;
Matches 43; Conservative 36; Mismatches 77; Indels 24; Gaps 7;
QY 87 TANLHLYDQLOREVQ-SLKEINFMLSVLQKEFLHLSKEFATTSKDLNVAQVDFYSCLOGF 145
DB 160 TQALESEQEKYNDTAQSLRDVTAQLESEQEKYNDTAQSLRDVTAQLESEQEKYNDTAQSL 219
QY 146 RDNVKGESLLDEYKNTST---EEMR--KLFSEIADLKGVSASLRREIRFLTPLAEEV 199
DB 220 RD-----VSAQLESYKSSSTLKEIEDLKLENTLQEKVAMAEKSVEDVQOQILTAESTNQEY 275
QY 200 RRLAHN-QESLTAIEELKTIKRDLSRDEIGQL-SQLSKTLTSOTALQKESSDLCQSIRE 257
DB 276 ARMVQDLQNRSLTKEEIKETITSSFLKIDTLKNOL-----RQODEDFRKQLEE 324

RESULT 11
US-08-700-178-3
Sequence 3, Application US/08700178
Patent No. 5783669
GENERAL INFORMATION:
APPLICANT: Hawkins, Phillip R.
APPLICANT: Wilde, Craig G.
APPLICANT: Seilhamer, Jeffrey J.
TITLE OF INVENTION: HYALURONAN RECEPTOR EXPRESSED IN HUMAN
TITLE OF INVENTION: UMBILICAL VEIN ENDOTHELIAL CELLS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08700,178
FILING DATE: August 20, 1996
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/402,217
FILING DATE: March 10, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0028-1 DIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
IMMEDIATE SOURCE:
LIBRARY: mouse
CLONE: GI 53979
US-08-700-178-3

Query Match 8.0%; Score 106; DB 1; Length 477;
Best Local Similarity 23.9%; Pred. No. 0.1;
Matches 43; Conservative 36; Mismatches 77; Indels 24; Gaps 7;
QY 87 TANLHLYDQLOREVQ-SLKEINFMLSVLQKEFLHLSKEFATTSKDLNVAQVDFYSCLOGF 145
DB 160 TQALESEQEKYNDTAQSLRDVTAQLESEQEKYNDTAQSLRDVTAQLESEQEKYNDTAQSL 219
QY 146 RDNVKGESLLDEYKNTST---EEMR--KLFSEIADLKGVSASLRREIRFLTPLAEEV 199
DB 220 RD-----VSAQLESYKSSSTLKEIEDLKLENTLQEKVAMAEKSVEDVQOQILTAESTNQEY 275
QY 200 RRLAHN-QESLTAIEELKTIKRDLSRDEIGQL-SQLSKTLTSOTALQKESSDLCQSIRE 257
DB 276 ARMVQDLQNRSLTKEEIKETITSSFLKIDTLKNOL-----RQODEDFRKQLEE 324

RESULT 12
US-08-995-654-3
Sequence 3, Application US/08995654
Patent No. 6025138
GENERAL INFORMATION:
APPLICANT: Hawkins, Phillip R.
APPLICANT: Wilde, Craig
APPLICANT: Seilhamer, Jeffrey
TITLE OF INVENTION: HYALURONAN RECEPTOR EXPRESSED IN HUMAN
TITLE OF INVENTION: UMBILICAL VEIN ENDOTHELIAL CELLS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/995,654
FILING DATE: December 22, 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/700,178
FILING DATE: August 20, 1996
CURRENT APPLICATION DATA:
APPLICATION NUMBER: 08/402,217
FILING DATE: March 10, 1995

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0028-2 DIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-895-0355
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-995-654-3

Query Match	8.0%	Score 106;	DB 3;	Length 477;
Best Local Similarity	23.9%	Pred. No. 0.1;		
Matches 43;	Conservative 36;	Mismatches 77;	Indels 24;	Gaps 7;

87	TANLHYODLQREVG-SLKEINFMLSVLQKEFLHLSKEFATTSKDLGSAVSQDFYSCLOGF	145
	: : : : : : : : : : : :	
Db		
160	TAQLESEGEKYNDAQSRQVTAQLESEGEKYNDTAQSLRDVTAQLESYQEKYNDTAQSL	219
QY		
146	RNYGFGFSLLDYKNST-----EEMR--KLFSQBIITADLKGSVASLREIRETLPLAEIV	199
	: : : : : : : : : : : :	
Db		
220	RD-----VSAQLESYKSTLKEIEDLKLENTLQEKYMAEKSVEDVQQQILTAESTNQEY	275
QY		
200	RLAHN-DESTAAIEELKTIKRSRDIEQCL-SQLSKILTQIALQRKESDLCQSIRE	257
	: : : : : : : : : : : :	
Db		
276	ARMVDQLNRTSLKEEIEKITSFBLKIDPLKNOL-----RQDDDFRKLQEE	324

```

RESULT 13
US-08-468-576B-12
: Sequence 12, Application US/08468576B
: Patent No. 5955345
: GENERAL INFORMATION:
: APPLICANT: Rabln, Daniel
: TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
: TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Spring Kramer Schaefer & Briscoe
: STREET: 660 White Plains Road
: CITY: Tarrytown
: STATE: New York
: COUNTRY: USA
: ZIP: 10591-5144
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
: COMPUTER: Apple Macintosh
: OPERATING SYSTEM: System 7.5
: SOFTWARE: WordPerfect
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/468,576B
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/239,276
: FILING DATE: 05-MAY-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/872,646
: FILING DATE: 08-JUN-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/715,181
: FILING DATE: 14-JUN-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/441,703
: FILING DATE: 04-DEC-1989
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/312,543

```

, FILING DATE: 17-FEB-1989
 , ATTORNEY/AGENT INFORMATION:
 , NAME: Kurt G. Briscoe
 , REGISTRATION NUMBER: 33,141
 , REFERENCE/DOCKET NUMBER: MDI 251.7-KGB
 , TELECOMMUNICATION INFORMATION:
 , TELEPHONE: (914) 332-1700
 , TELEFAX: (914) 332-1844
 , INFORMATION FOR SEQ ID NO: 12:
 , SEQUENCE CHARACTERISTICS:
 , LENGTH: 604 amino acids
 , TYPE: amino acid
 , TOPOLOGY: linear
 US-08-468-576B-12

Query Match	8.0%	Score 106;	DB 2;	Length 604;
Best Local Similarity	25.2%;	Pred. No. 0.14;		
Matches	38;	Conservative 37;	Mismatches 48;	Indels 28; Gaps 5;
QY	97	QREVGSLKEINFMLSVLQKSEFLHLSKEFATSKDLSAVSQDFYSCLOGFRDNYKGFESLL	156	
Db	121	ERYKGSUADV--VDLUKQKLL-----EEMTRTEQEDSSCM-----BKLL	157	
QY	157	DEYKNSTEEMRKLFQSBIIADLKGSVASLREEIRFLPLAEAEVRLAHNOESLTAATEEL	216	
Db	158	S--KDWKEKMERLNTSELLGEIKGTPSLAEKERQLSTM---ITQLISLREQLLAAHDEQ	212	
QY	217	KTRDSLRDDEIGQLSQSLTYLTTSQIALQORKE	247	
Db	213	KKLAAQIEKIQKQRMQMDLAQQQEQIARQQQQ	243	

```

RESULT 14
US-08-468-579B-12
; Sequence 12, Application US/08468579B
; Patent No. 5981700
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,579B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,276
; FILING DATE: 05-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,646
; FILING DATE: 08-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 17-FEB-1989

```

Search completed: August 13, 2002, 09:24:21
Job time: 152 sec

```

1 RESULT 15
2 US-08-468-577B-12
3 ; Sequence 12, Application US/08468577B
4 ; Patent No. 6001804
5 ; GENERAL INFORMATION:
6 ; APPLICANT: Rabin, Daniel
7 ; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
8 ; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
9 ; NUMBER OF SEQUENCES: 19
10 ; CORRESPONDENCE ADDRESS:
11 ; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
12 ; STREET: 660 White Plains Road
13 ; CITY: Tarrytown
14 ; STATE: New York
15 ; COUNTRY: USA
16 ; ZIP: 10591-5144
17 ; COMPUTER READABLE FORM:
18 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
19 ; COMPUTER: Apple Macintosh
20 ; OPERATING SYSTEM: System 7.5
21 ; SOFTWARE: WordPerfect
22 ; CURRENT APPLICATION DATA:
23 ; APPLICATION NUMBER: US/08/468.577B
24 ; FILING DATE: 06-JUN-1995
25 ; CLASSIFICATION: 514
26 ; PRIOR APPLICATION DATA:
27 ; APPLICATION NUMBER: US 08/239,276
28 ; FILING DATE: 05-MAY-1994
29 ; PRIOR APPLICATION DATA:
30 ; APPLICATION NUMBER: US 07/872,646
31 ; FILING DATE: 08-JUN-1992
32 ; PRIOR APPLICATION DATA:
33 ; APPLICATION NUMBER: US 07/715,181
34 ; FILING DATE: 14-JUN-1991
35 ; PRIOR APPLICATION DATA:
36 ; APPLICATION NUMBER: US 07/441,703
37 ; FILING DATE: 04-DEC-1989
38 ; PRIOR APPLICATION DATA:
39 ; APPLICATION NUMBER: US 07/312,543
40 ; FILING DATE: 17-FEB-1989
41 ; ATTORNEY/AGENT INFORMATION:

```


	Query Match	12.2% score 162.5; DB 2;	Length 355;
	Best Local Similarity	21.4%; pred. No. 0.0025;	
Matches	63; Conservative	61; Mismatches 111;	Indels 59; Gaps
QY	8 VTTPPPAPSYSANRY-----PQSLMDKIKTKIAATASLLICTICFALLGHLGV 58	:	:
	:	: : :	: :
DB	32 IIPISIEAPTSSAAVAGAKTAIEBGRSPFLQRICYLVKIITAIATLFWVGIALVCLYLGV 91	:	:
	: : :	: : :	: :
QY	59 FLIQAOTTIVLLALFITSLA-----GNALYLQKTANLHLYODLOR-----EV 100	:	:
	: : :	: : :	: :
DB	92 SVISTPSLLMLAIMLVSEFVITAIRDGTPTSVVR---HMKKIQOQGEEINTRLHTAV 147	:	:
	: : :	: : :	: :
QY	101 GSUKETNFMLS-VLOKEFLH-----LSKEPATTSKDLSVASQDFYSCLQGFDRNYKGFE 153	:	:
	: : :	: : :	: :
DB	148 ENLKAYNVELSEGINOLKQHTRLSDFGDRLEANTGDFTALIADPOLSLEEFKSVMGKVE 207	:	:
	: : :	: : :	: :
QY	154 SILDEYKNSTTEEMRKLFSGEIIADIADLGVSASLRSEIRFLTPLAEVRRLAHNQESLTAAI 213	:	:

Qy	3	TPTLVIP-PSPAPSYSANRVPQPSIMDKIKKIAIASLIIIGTIGFIALGLHGVGF---	59
Db	8	TPSAPNPIPAPTGPITPK-PRSF---IEKVIIVAKYLFAIANTSGALGTILGSG	63
Qy	60	LIAPOITVILLAFITSLAGNALYLOQTANLHLHYQDLQREVSKLEINMFLSVLOKEFLH	119
Db	64	ALTGIGIALLVIFFVMVLGLIKDSISGGERRLRREEVSRFTSENORLIV-----	116
Qy	120	LSKEFATTSKDLSAVSDFYSCLOGFRD---NVK-----	150
Db	117	ITTTLETVEVDLKAAKDOQLILEEAFNENGNLKTAEDEEQVSKLSEOLEALERINOL	176
Qy	151	-----GFES-LUDEYKNSTEEMRKLFSEOIIDADLGKSWASLREEI	189
Db	177	IOANAGDAQETSSELKKLIISGWDSKVVEQNTSIQALKVLLGOEWOEQAETHVKAMQEQI	236
Qy	190	RFL-----TPLAEEVRRLAHNOESILITAEIELKTIIRD-----SLRDEIQOLS	231
Db	237	QALQAEILGMHNQSTALOXSVENLLVODQALTRVVGELLESENKLISOACSALRQEIETKLA	296
Qy	232	QLSKTLTSQI-----ALQR-----KSSDLCSQIRETLSSPPKSAASPST	270
Db	297	OHEITSLOORIDAMLQAQNLEAOVTALEKKMOEAKOEAESFEICAFVDRRTTCGRETPPPTT	356

RESULT 5
B86514 hypothetical protein CPJ0186 [imported] - Chlamydophila pneumoniae (strain J138)
C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: B86514
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba

[illegible]

85 QKTA^{NH}LYQD^LQREVGSLKEIN^FMLSVLQKEFLHLSKEFATTSKDL^SAVSQDYFSC

Db 668 EKVENLRIMKESLEGEVNSLR---VKLKALENQSPFLRIRMSDVEKEISLSIKDLEKLI- 723

QY 144 GFRDNTYKGFESLLDEYKNTTEMRKL-----FSQEIADLKGSVASL----- 185

Db 724 -----KEBESLRSEIEDSERKIAEIDETISKKDQVAKLKGRIERLKRDRDKLKALEN 777

QY 186 -----REIRFLTPLAEVRRLAHNOESLTAAI--EELKTIKIRDSLRDEI--GOLSQ 233

Db 778 PEAREVTEKIREVEREIAKREELSVEGKLESLSRLNDELIPRKASLEEEIEGLVNI 837

QY 234 S-----KTLTSQIALQKESDLCQIRE 257

Db 838 NALKANINENEALKSLTEKLEKLEKKEGEIYSRIEE 874

RESULT 9

S53348

Alternate names: protein YCLO29c

Species: Saccharomyces cerevisiae

R;Variety: strain carlsbergensis

C;Date: 06-Jun-1995 #sequence_revision 21-Jul-1995 #text_change 06-Feb-1998

C;Accession: S53348; S70387

R;Porter, G.L.

Submitted to the EMBL Data Library, August 1994

A;Description: The polyfunctional HIS4 gene of Saccharomyces carlsbergensis: sequence an

A;Reference number: S53348

A;Accession: S53348

A;Molecule type: DNA

A;Residues: 1-347 <POR>

A;Cross-references: EMBL:U13062; NID:g532779; PID:g532780

R;Porter, G.; Westmoreland, J.; Priebe, S.; Resnick, M.A.

Genetics 143, 755-767, 1996

A;Title: Homologous and homeologous intermolecular gene conversion are not differential

nd MSH2.

A;Reference number: S70387; MUID:96363910

A;Accession: S70387

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-347 <POW>

A;Cross-references: EMBL:U13062; NID:g532779; PID:g532780

A;Note: the source is designated as Saccharomyces carlsbergensis

C;Genetics:

A;Gene: SGD:B1K1

A;Cross-references: SGD:S0000534; MIPS:YCL029c

A;Map position: 3L

Query Match 9.3%; Score 124; DB 2; Length 347;

Best Local Similarity 27.1%; Pred. No. 0.72;

Matches 56; Conservative 25; Mismatches 86; Indels 40; Gaps 8;

QY 86 KLANHLQYDLOREVG-SLKEINFMLSVLOKEFLHLSKEFATTSKLSAVSQDFYSCIQ 144

Db 86 KASNSHVNDHNSVETIELRAQLTIKLRQLHYKRLDDQRMVLEEQPTF----- 139

QY 145 FRDNTYKGF-----ESLLDEYKNTSE-EMRKLPFSQEIAD-----LKGVSASLRREIRLTP 194

Db 140 --DYEATIQDREREIDELKQOLEHHRQAKOFFDAENEQQLQAVVSQLEHEIK----- 193

QY 195 LAEEVRLAHNQ-----ESLTAIEELKTIKIRDSLRDEIGOLSQSLKTLTSQIALQ 246

Db 194 -ENEERNIAHNOPTDSSIEDIELKKQVQLRSIEDQF-----ELHKTWKAREQLK 244

QY 247 ESSDLCQSIRETTLSPRKSASPSTKSS 273

Db 245 MHNDSLSRKYONLSKELSTKPDQSSS 271

RESULT 10

A56539

giantin - human

N;Alternate names: macrogolglin

C;Species: Homo sapiens (man)

C;Date: 19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change 10-Dec-1999

C;Accession: A56539; S37536

R;Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M.

Mol. Cell. Biol. 14, 2564-2576, 1994

A;Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protei

A;Reference number: A56539; MUID:94187728

A;Accession: A56539

A;Molecule type: mRNA

A;Residues: 1-3259 <SEE>

A;Cross-references: EMBL:X75304; NID:g405714; PIDN:CAA53052.1; PID:g405715

C;Genetics:

A;Gene: GDB:GOLGB1; GCP; GCP371

A;Cross-references: GDB:454958

A;Map position: 3q13.31-3q13.31

C;Superfamily: giantin

C;Keywords: coiled coil; Golgi apparatus; transmembrane protein

F:3238-3254/Domain: transmembrane #status predicted <TMN>

Query Match 9.2%; Score 122.5; DB 1; Length 3259;

Best Local Similarity 23.6%; Pred. No. 13;

Matches 53; Conservative 35; Mismatches 90; Indels 47; Gaps 7;

QY 94 QDLQREVGLKEINFMLSV-----LQKFLHLSKEFATTSKLSAV-----S 135

Db 1584 EKLKVEISLKSKEIAESTEWQEKHELOKEYEILLQSYENVSNSEARIQHVVEAVRQEK 1643

QY 136 QDFYSCLOGFRDNYKGFESLLDEYKNTSEMRKL-----SQEIIADLKGSVASLRREI 189

Db 1644 QELYGKLRSKTEANKKTEKQLOAEQEMEEKMKKFAKSKQKILELEENDRLRAEV 1703

QY 190 R-----FLPLAEVRLAHNQESLTAIEELKTIKIRDSLRDEIGOLS-OL 233

Db 1704 HPAGDTAKCMTLLSSNASMKELERVMEXETLSKTKQSLMSEKDSUSEEVQDLKHQI 1763

QY 234 SKTLTSQIALQKESDLCQIRE--TLSSP-----RKGSASPSTK 271

Db 1764 EDNVSKQANLEATEKHNDQNTVTEGTQSIPTGETEQDSLSMSTR 1808

RESULT 11

VPXRW8

N;Alternate names: hemagglutinin

N;Contains: outer capsid protein VP5; outer capsid protein VP8

C;Species: human rotavirus A

C;Date: 30-Jun-1990 #sequence_revision 30-Jun-1990 #text_change 01-Mar-1996

C;Accession: F28839

R;Gorziglia, M.; Green, K.; Nishikawa, K.; Taniguchi, K.; Jones, R.; Kapikian, A.2.;

J. Virol. 62, 2978-2984, 1988

A;Title: Sequence of the fourth gene of human rotaviruses recovered from asymptomatic

A;Reference number: A28839; MUID:88275070

A;Accession: F28839

A;Molecule type: genomic RNA

A;Residues: 1-775 <GOR>

C;Genetics:

A;Map position: segment 4

C;Superfamily: rotavirus outer layer protein VP3

C;Keywords: capsid protein; coat protein; glycoprotein; hemagglutinin

F:1-240/Product: outer capsid protein VP8 #status predicted <VP>

F:241-246/Region: cleavage processing #status predicted

F:247-775/Product: outer capsid protein VP5 #status predicted <VP5>

F:32,56,85,97,111,114,132,192,277,324,583,606/Binding site: carbohydrate (Asn) (coval

Query Match 9.2%; Score 122; DB 1; Length 775;

Best Local Similarity 26.8%; Pred. No. 2.5;

Matches 56; Conservative 33; Mismatches 86; Indels 34; Gaps 9;

QY 73 FITSLAGNALYQKTAN-LHLYQDLQREVGLSKELNFMLSVLOKEFLHLSKEFATTSK-D 130

Search completed: August 13, 2002, 09:24:56
Job time: 152 sec

GenCore version 4.5

Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 09:30:42 ; Search time 17.41 Seconds
(without alignments)
607.147 Million cell updates/sec

Title: US-09-673-763-14

Perfect score: 1333

Sequence: 1 MTPTTLIVPPSPAPSYSA.....QIRETLSPKSPSTKSS 273

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	128.5	9.6	1957	1 YD86_SCHPO	Q10411 schizosacch
2	124.5	9.3	3210	1 CENF_HUMAN	P49454 homo sapien
3	122	9.2	775	1 VP4_ROTTH1	P11198 human rotav
4	120	9.0	496	1 Y7J3_ANASP	Q05070 anabaena sp
5	116	8.7	776	1 VP4_ROTTHN	P11199 human rotav
6	115	8.6	1325	1 G160_MOUSE	P55937 mus musculus
7	114	8.6	775	1 VP4_ROTTHG	P23045 porcine rot
8	112.5	8.4	622	1 LAM0_DROME	P08928 drosophila
9	111.5	8.4	391	1 APA4_RAT	P02651 rattus norv
10	111.5	8.4	775	1 VP4_ROTTHM	P11197 human rotav
11	111.5	8.4	1044	1 YAF3_SCHPO	Q09857 schizosacch
12	111.5	8.4	1790	1 US01_YEAST	P25386 saccharomyc
13	111	8.3	432	1 GRAP_HUMAN	P14136 homo sapien
14	110.5	8.3	2104	1 MY53_SCHPO	O14157 schizosacch
15	110	8.3	774	1 VP4_ROTTH	P11200 human rotav
16	110	8.3	978	1 RA50_AQUAE	O67124 aquifex aeo
17	109.5	8.2	776	1 ALM1_SCHPO	Q08778 human rotav
18	109.5	8.2	1727	1 OV71_ONCVO	Q9utK5 schizosacch
19	109	8.2	432	1 VP4_ROTTH	P31732 onchocerca
20	109	8.2	776	1 TPR_HUMAN	Q06894 canine rota
21	109	8.2	2349	1 K2C8_HUMAN	P12270 homo sapien
22	108.5	8.1	1453	1 Y373_BOVIN	P05787 homo sapien
23	108.5	8.1	504	1 AINX_MOUSE	Q9tu23 bos taurus
24	108	8.1	804	1 MEA6_HUMAN	P46660 mus musculus
25	108	8.1	1251	1 CYA8_HUMAN	O15320 homo sapien
26	108	8.1	804	1 AINX_MOUSE	P40145 homo sapien
27	107.5	8.1	1005	1 RA50_METJA	Q58718 methanococc
28	107	8.0	535	1 HTR1_HALNI	P33741 halobacteri
29	107	8.0	776	1 VP4_ROTTH	Q07416 feline rota
30	107	8.0	776	1 VP4_ROTTH	P08729 human rotav
31	106.5	8.0	468	1 K2C7_HUMAN	Q06895 human rotav
32	106.5	8.0	729	1 KAR3_YEAST	P17119 saccharomyc
33	106.5	8.0	1189	1 SCII_CHICK	Q90988 gallus gall

34	106.5	8.0	2663	1 CENE_HUMAN	Q02224 homo sapien
35	106	8.0	287	1 APAL_MACFA	P15568 macaca fasc
36	106	8.0	499	1 AINX_HUMAN	Q16352 homo sapien
37	106	8.0	505	1 AINX_RAT	P23565 rattus norv
38	106	8.0	775	1 VP4_ROTTH3	P39033 human rotav
39	106	8.0	776	1 VP4_ROTTH1	P12976 simian 11 r
40	106	8.0	776	1 VP4_ROTTH1	P17464 simian 11 r
41	106	8.0	794	1 HMR_MOUSE	Q00347 mus musculus
42	106	8.0	827	1 SOX6_MOUSE	P40645 mus musculus
43	106	8.0	882	1 MYSP_CAEEL	P10567 caenorhabdi
44	105.5	7.9	775	1 VP4_ROTTH5	P30214 human rotav
45	105.5	7.9	775	1 VP4_ROTTHD	P11196 human rotav

ALIGNMENTS

RESULT 1
YD86_SCHPO
ID YD86_SCHPO STANDARD; PRT; 1957 AA.
AC Q10411;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 222.8 kDa protein ClpF3.06C in chromosome I.
GN SPAC1F3.06C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Connor R., Churcher C.M., Bartell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 270690; CAA94624.1; -
KW Hypothetical protein.
SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;

Query Match 9.68; Score 128.5; DB 1; Length 1957;
Best Local Similarity 27.48; Pred. No. 1;
Matches 51; Conservative 32; Mismatches 82; Indels 23; Gaps 5;
QY 89 NLHYLDQREVGSLSKEINFLMSVLQKEFLHLSKEFATTSKLDLSAVSQDFYSCIQGFQDN 148
DB 345 NEKLEKLLRNTIGSLKDSRTSNLSOLEEEMVELKESNRTIHSQLT-----DAESKLSFQE 400
QY 149 YGFESLLDEYKKNSTEMRKLFSQETIADLKGVSASLREIRFLTPLAEEVRLAH---N 205
DB 401 NKSLSGSDIYQNNLSKDKMKVKQ-----VSSQLEEAR--SSLAHATGKLAETNSE 449
QY 206 QESLTAAEELKTRISRLDEIGQLSQTLSQALQKESSDLCSTIRETLSPKRS 265
DB 450 RFOQNKIKDFKIEQDLRACLNSNNELKESALDKKKDQLNNLRQIKE-----QKK 504
QY 266 ASPSTKSS 273
DB 505 VSESTOSS 512

RESULT 2
CENF_HUMAN

ID AC P49454; Q13246; Q13171; PRT; 3210 AA.
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE CENP-F kinetochore protein (Centromere protein F) (Mitosis) (AH
 DE antigen).
 GN CENPF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Breast Carcinoma;
 RX MEDLINE=95348175; PubMed=7542657;
 RA Liao H., Winkler R.J., Mack G., Rattner J.B., Yen T.J.;
 RT "CENP-F is a protein of the nuclear matrix that assembles onto
 RT kinetochores at late G2 and is rapidly degraded after mitosis.";
 RT J. Cell Biol. 130:507-518(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95379848; PubMed=7651420;
 RA Zhu X., Mancini M.A., Chang K.-H., Liu C.-Y., Chen C.-F., Shan B.,
 RA Jones D., Yang-Peng T.H., Lee W.-H.;
 RT "Characterization of a novel 350-kilodalton nuclear phosphoprotein
 RT that is specifically involved in mitotic-phase progression.";
 RL Mol. Cell. Biol. 15:5017-5029(1995).
 RN [3]
 RP SEQUENCE OF 2194-3210 FROM N.A.
 RX MEDLINE=95336446; PubMed=7612011;
 RA Li Q., Ke Y., Kapp J.A., Fertig N., Medsger T.A. Jr., Joshi H.C.;
 RT "A novel cell-cycle-dependent 350-kDa nuclear protein: C-terminal
 RT domain sufficient for nuclear localization.";
 RL Biochem. Biophys. Res. Commun. 212:220-228(1995).
 RN [4]
 RP CHARACTERIZATION.
 RX MEDLINE=95370296; PubMed=7642639;
 RA Zhu X., Chang K.-H., He D., Mancini M.A., Brinkley W.R., Lee W.-H.;
 RT "The C terminus of mitosis is essential for its nuclear localization,
 RT centromere/kinetochore targeting, and dimerization.";
 RL J. Biol. Chem. 270:19545-19550(1995).
 RN [5]
 RP CHARACTERIZATION.
 RX MEDLINE=98437347; PubMed=9763420;
 RA Chan G.K.T., Schaar B.T., Yen T.J.;
 RT "Characterization of the kinetochore proteins CENP-F and hBUBR1.";
 CC Interactions with the kinetochore proteins CENP-F and hBUBR1.";
 CC J. Cell Biol. 143:49-63(1998).
 CC -!- FUNCTION: PROBABLY REQUIRED FOR KINETOCHORE FUNCTION, INVOLVED IN
 CC CHROMOSOME SEGREGATION DURING MITOSIS. INTERACTS WITH
 CC RETINOBLASTOMA PROTEIN (RB), CENP-E AND BUBR1.
 CC -!- SUBUNIT: HOMO- OR HETERODIMER.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR MATRIX (BUT NOT IN THE NUCLEOLUS),
 CC REORGANIZATION TO THE KINETOCHORE/CENTROMERE (CORONAL SURFACE OF
 CC THE OUTER PLATE) AND THE SPINDLE DURING MITOSIS.
 CC -!- DEVELOPMENTAL STAGE: GRADUALLY ACCUMULATES DURING THE CELL CYCLE.
 CC -!- PTM: HYPERPHOSPHORYLATED DURING MITOSIS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; U19769; AAA82889.1; .
 CC EMBL; U30872; AAA82935.1; .
 CC EMBL; U25725; AAA86889.1; .
 CC HSP; P02649; ILE4.
 CC MIM; 600236; .
 CC
 CC Chromosomal protein; Nuclear protein; Centromere; Coiled coil;
 KW
 Mitosis; Phosphorylation; Antigen; Cell cycle; Repeat.
 KW DOMAIN 14 197
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 273 769
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 823 1328
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1642 1746
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1862 1987
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 2207 2568
 FT COILED COIL (POTENTIAL).
 FT REPEAT 2386
 FT REPEAT 2389
 FT REPEAT 3015
 FT REPEAT 3032
 FT DOMAIN 16
 FT CONFLICT 250 250
 FT CONFLICT 272 272
 FT CONFLICT 611 611
 FT CONFLICT 1494 1589
 FT CONFLICT 1611 1611
 FT CONFLICT 1811 1811
 FT CONFLICT 2242 2243
 FT CONFLICT 2335 2335
 FT CONFLICT 2492 2492
 FT CONFLICT 2545 2545
 FT CONFLICT 3210 AA; 367589 MW; 11D83324960E4334 CRC64;
 SQ
 Query Match 9.3%; Score 124.5; DB 1; Length 3210;
 Best Local Similarity 26.2%; Pred. No. 3.4;
 Matches 56; Conservative 32; Mismatches 83; Indels 43; Gaps 8;
 QY 77 LAGNALYLOK-----TANLHLYQDLOREVGSLKEINFMLSVLQKEFHLHLSKEFAT--- 126
 Db 946 LLETSLSLEKEMSSIIISLNKREIELTQENGLKEINASLNQKMNLIQKSESFANYID 1005
 QY 127 -TSKDLGAVS----QDFYSLQGFQFRNYKGFESLLDYKNSTEEMRK-----FSEIITA 176
 Db 1006 EREKSISELSDYKQKELLILQRCETGNAYEDLSQYKAAQKSKLECLNLECTSLICE 1065
 QY 177 DLKGSVASLREET----RFLTPLAEVRLAHNQSLTAAIEELTIRDSLRDEI---- 227
 Db 1066 NRKNELEQLKEAKEHKEHLETLKLAFAER---NQNLML-----ELETVQOALRSEMTDNO 1118
 QY 228 -----GQLSLSKLTLSQIALQKSESSDLCSQ 254
 Db 1119 NNSKSEAGLQKQELMTLKEQNKMQKEVNDLQEQ 1152
 RESULT 3
 ID VP4_ROT1 STANDARD; PRT; 775 AA.
 AC P11198;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer capsid protein VP4 (Hemagglutinin) (Outer layer protein VP4)
 DE [Contains: Outer capsid proteins VP5 and VP8].
 GN S4.
 OS Human rotavirus (serotype 1 / strain 1076).
 OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.
 OX NCBI_TaxID=10944;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86275070; PubMed=2839714;
 RA Gorziglia M., Green K., Nishikawa K., Taniguchi K., Jones R.,
 RA Kapikian A.Z., Chanock R.M.;
 RT "Sequence of the fourth gene of human rotaviruses recovered from
 RT asymptomatic or symptomatic infections.";
 RL J. Virol. 62:2978-2984(1988).
 RN [2]
 RP SEQUENCE OF 1-280.
 RX MEDLINE=86313706; PubMed=3018754;
 RA Gorziglia M., Hoshino Y., Buckler-White A., Blumentals I., Glass R.,
 RA Flores J., Kapikian A.Z., Chanock R.M.;
 RT "Conservation of amino acid sequence of VP8 and cleavage region of

RT 84-Kda outer capsid protein among rotaviruses recovered from
 RL asymptomatic neonatal infection.;
 CC Proc. Natl. Acad. Sci. U.S.A. 83:7039-7043(1986).
 CC -1- SUBCELLULAR LOCATION: Outer capsid.
 CC -1- PTH: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE
 CC OTHER PRODUCT IS VP5.
 CC -1- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.
 DR PIR: F28839; VPXRW8.
 DR PIR: F25904; VPXR16.
 DR InterPro: IPR00416; Cap_VP4.
 DR Pfam: PF00426; VP4; 1.
 KW Coat protein; Glycoprotein.
 FT CHAIN 1 775
 FT CHAIN 1 240
 FT CHAIN 247 775
 FT CARBOHYD 32 56
 FT CARBOHYD 56 32
 FT CARBOHYD 97 97
 FT CARBOHYD 111 111
 FT CARBOHYD 114 114
 FT CARBOHYD 132 132
 FT CARBOHYD 192 192
 FT CARBOHYD 277 277
 FT CARBOHYD 324 324
 FT CARBOHYD 583 583
 FT CARBOHYD 606 606
 SQ SEQUENCE 775 AA; 87697 MW; EE29A36D36C8018F CRC64;

Query Match 9.2%; Score 122; DB 1; Length 775;
 Best Local Similarity 26.8%; Pred. No. 0.92; Mismatches 33; Conservative 33; Indels 86; Gaps 9;
 Matches 56; Conservative 33; Mismatches 86; Indels 34; Gaps 9;
 QY 73 FITSLAGNALYLQKTN-LHLQDLOREVGLSKENFMSLVQKEFLHLSKEFATSK-D 130
 Db 469 FILLVPSNDYDTPIMNSVTVRQDLERQLGLRE-----EFNSLQEIATQLID 518
 QY 131 LSAVSQDFYSCLOQGRDNYKGFESLDKYNSTDEMRKLFQEIADLKGVSASIREIR 190
 Db 519 LALLPLDMFS-----FSGIKSIDAKSNATWVMKFKK-----SGLSATSEISLTR 565
 QY 191 FLTPLAEEVRR---LAHQESL---TAAIEELTIROSLRDEIGQLSLSKTL-TSIAL 243
 Db 566 SILSNAASSVRSKSIKSNISVWTVDSQITGSSDSVRNISTQTSASKRLRLRIIT 625
 QY 244 QRKES--DLCSQIRETLSSPKRSASPST 270
 Db 626 QTEGMNFIDISAALVKTIDKSTHISPT 654

RESULT 4
 ID V7J3_ANASP STANDARD; PRT; 496 AA.
 AC Q05070;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein alr4393 precursor.
 GN ALR4393.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN [1]
 RP MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Iriuguchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 cyanobacterium Anabaena sp. strain PCC 7120.";
 RL DNA Res. 8:205-213(2001).

RP SEQUENCE OF 1-166 FROM N.A.
 RX MEDLINE=93308081; PubMed=8391534;
 RA Wei T.-F., Ramasubramanian T.S., Pu F., Golden J.W.;
 RT "Anabaena sp. strain PCC 7120 difa gene encoding a sequence-specific
 RT DNA-binding protein cloned by in vivo transcriptional interference
 RT selection.";
 RL J. Bacteriol. 175:4025-4035(1993).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AP003596; BAB76092.1; -
 DR EMBL; L10036; AAD04182.1; -
 KW Hypothetical protein; Transmembrane; Signal; Complete proteome.
 FT CHAIN 1 19 POTENTIAL.
 FT SIGNAL 1 19 POTENTIAL.
 FT TRANSMEM 20 496 HYPOTHETICAL PROTEIN ALR4393.
 FT TRANSMEM 45 67 POTENTIAL.
 SQ SEQUENCE 496 AA; 56375 MW; 454D67CF420AFB9E CRC64;
 Query Match 9.0%; Score 120; DB 1; Length 496;
 Best Local Similarity 21.9%; Pred. No. 0.73; Mismatches 57; Conservative 54; Mismatches 87; Indels 62; Gaps 11;
 Matches 57; Conservative 54; Mismatches 87; Indels 62; Gaps 11;
 QY 38 IASILITGTTGFLALLGLHVG-----FLIAPQITIVLLALFTSLAGNALYLQKT 87
 Db 8 IAILILG--GVATVGDRIQTRVGRKARLSLFLNPKNTAVL-----VTILTGLVSATTL 61
 QY 88 ANLHL-----YQDLQREVGLSKENFMSLVQK-----EFLHLSKEFATTSK 129
 Db 62 AILFIADDEGLRGVFELEDIQDLQKRE---OLKVAEEQKQVEIERKNVNOLETTT 118
 QY 130 DLSAV-----SQDFYSCLOQGRDNYKGFESLDKYNSTDEMRKLFQEI 175
 Db 119 DKQVETQRDQAKKEKLKAQDLAQAOYQRTQSLGOVTVQYQKAIELASVNR-- 176
 QY 176 ADLKGVSASIREIRELTPLAEV-----RLAHQESLTAIEELKTIROSLRDEIGQL 230
 Db 177 KALQGAVEQLKTERRLYAEAKKAIEQRDLRANRQQAIEQRDLRANRQQAIEQRDLR 236
 QY 231 SOLSKTLTS---QIALQKRE 247
 Db 237 SOLDKIIQNRNLEIA-QREE 255

RESULT 5
 ID VP4_ROTNN STANDARD; PRT; 776 AA.
 AC P11199;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer capsid protein vp4 (Hemagglutinin) (Outer layer protein VP4)
 DE [Contains: Outer capsid proteins VP5 and VP8].
 GN S4.
 OS Human rotavirus (serotype 3 / strain McN13).
 OC Viruses; dsRNA viruses; Reoviridae; Rotaviruses.
 OX NCBI_TaxID=10955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88275070; PubMed=2839714;
 RA Gorziglia M., Green K., Nishikawa K., Taniguchi K., Jones R.,
 RA Kapikian A.Z., Chanock R.M.;
 RT "Sequence of the fourth gene of human rotaviruses recovered from
 RT asymptomatic or symptomatic infections.";
 RL J. Virol. 62:2978-2984(1988).
 RN [2]

Tue Aug 13 14:23:41 2002

RA Kondo M., Sutou S.;
RT "Cloning and molecular characterization of cDNA encoding a mouse
RT male-enhanced antigen-2 (Mea-2): a putative family of the Golgi
RL autoantigen.";
RL DNA Seq. 7:71-82(1997).
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN SPERMATOGENESIS AND/OR
CC TESTIS DEVELOPMENT. PROBABLY IDENTICAL WITH THE SEROLOGICALLY
CC DETECTABLE MALE ANTIGEN (SDM).
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTIS. TRANSCRIPTS CAN BE
CC FOUND IN SPERMATIDS DURING SPERMATOGENESIS. NO EXPRESSION IN
CC LEYDIG CELLS, SPERMATOGENIA, OR SPERMATOCYTES.
CC -!- SIMILARITY: HIGH, TO HUMAN GOLGIN-160.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER ME1-1, ME1-19 OR ME1-30 IS THE
CC INITIATOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D78270; BAA19612.1; -
DR HSSP; P18852; ISCG.
DR MGD; MGI:96958; Golga3.
KW Spermatogenesis; Developmental protein.
FT DOMAIN 201 204 POLY-ALA.
SQ SEQUENCE 1325 AA; 149880 MW; 3230636962C687B0 CRC64;

Query Match 8.6%; Score 115; DB 1; Length 1325;
Best Local Similarity 21.4%; Pred. No. 4-9;
Matches 48; Conservative 51; Mismatches 83; Indels 42; Gaps 8;

QY 68 VLLALFITSAGNALYLOKTAHLHYQDLQREVSLKEINFMLSVLQKEFLHLSKEFATT 127
DB 610 IVLEVALQSAKDEKELDRGAR-RLEEDTETSLGLLEQLQDLQAVKSNQVEHLQOETATL 668
QY 128 SKDLSANVSQDFYS---CLOQFRDNYKGFESLLDEYKKNSTE---EMRKLFSQEI----- 175
DB 669 RKQMKVKEQFVQOKVMVEAYRRDATSKDOLINELKATKRLDSEMKEL-ROELIKLOGE 727
QY 176 -----ADLKGVSASLREIRELFTPLAEVRLAINQESLTAAI 213
DB 728 KKTVEVHSRLQKMSLVHQMAELHGLQSVQKE---RDEWEIHLQSLKDFKDEQMIALT 784
QY 214 ELKTIKSLRDEIGQLSKLTSQIALQKESDLCQSIRE 257
DB 785 EANETLKQI-BELOO--EAKKAITEQKMKRLGSLDLSAQKE 825

RESULT 7
VP4_ROTPTG STANDARD; PRT; 775 AA.
ID VP4_ROTPTG
AC P23045;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer capsid protein VP4 (Hemagglutinin) (Outer layer protein VP4)
DE [Contains: Outer capsid proteins VP5 and VP8].
DE S4.
OS Porcine rotavirus (strain Gottfried).
OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX NCBI_TaxID=10917;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90080150; PubMed=2152826;
RA Gorziglia M., Nishikawa K., Hoshino Y., Taniguchi K.;
RT "Similarity of the outer capsid protein VP4 of the Gottfried strain
RT of porcine rotavirus to that of asymptomatic human rotavirus
RL J. Virol. 64:414-418(1990)."

RP SEQUENCE OF 1-281.
RX MEDLINE=86313706; PubMed=3018754;
RA Gorziglia M., Hoshino Y., Buckler-White A., Blumentals I., Glass R.,
RA Flores J., Kapikian A.Z., Chanock R.M.;
RT "Conservation of amino acid sequence of VP8 and cleavage region of
RT 84-kDa outer capsid protein among rotaviruses recovered from
RT asymptomatic neonatal infection."
RL Proc. Natl. Acad. Sci. U.S.A. 83:7039-7043(1986).
CC -!- SUBCELLULAR LOCATION: Outer capsid.
CC -!- PTM: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE
CC OTHER PRODUCT IS VP5.
CC -!- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.
CC
CC PIR: G28639; VPXRW9.
DR PIR: G25904; VPXRWN.
DR InterPro; IPR000416; Cap_VP4.
DR Pfam; PF00426; VP4; 1.
KW Coat protein; Glycoprotein.
FT CHAIN 1 776
FT CHAIN 248 776
FT CARBOHYD 32
FT CARBOHYD 56
FT CARBOHYD 97
FT CARBOHYD 111
FT CARBOHYD 114
FT CARBOHYD 132
FT CARBOHYD 133
FT CARBOHYD 193
FT CARBOHYD 278
FT CARBOHYD 325
FT CARBOHYD 584
FT CARBOHYD 607
FT CARBOHYD 607
FT CONFLICT 190 190
SQ SEQUENCE 776 AA; 87794 MW; CFCADF4DEE986512 CRC64;

Query Match 8.7%; Score 116; DB 1; Length 776;
Best Local Similarity 25.8%; Pred. No. 2,2;
Matches 48; Conservative 31; Mismatches 75; Indels 32; Gaps 7;

QY 94 QDLQREGLSLKEINFMLSVLQKEFLHLSKEFATSK-DLSANVSQDFYSCLQFRDNYKGF 152
DB 492 QDLERQLGLDRE-----EFNLSQEIATQILDALPLDMFMS-----FSGI 534
QY 153 ESSLDEYKKNSTEEMRKLFSOEIITADLKGVSASLREIRELFTPLAEVRR---LAHQESL 209
DB 535 KSTIDAAKSKATKMKKFKR-----SGLATSISELTRLSNAASSYSSRSIRSNISSI 588
QY 210 ---TAATEELKTIKSLRDEIGQLSKLTSQIALQKESDLCQSIRETLSSPRK 264
DB 589 SWTDVVSQITGSSDSVRNISTQTSASRRLRLREITTTQTEGMNFIDISAAVLTKIDKST 648

QY 265 SASPST 270
DB 649 HISPT 654

RESULT 6
G160_MOUSE STANDARD; PRT; 1325 AA.
ID G160_MOUSE
AC P55937;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Golgin-160 (Male-enhanced antigen-2) (MEA-2).
DE GOLGA3 OR MEA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CD-1; TISSUE=Testis;
RX MEDLINE=97217683; PubMed=9063644;

Db 210 TLRVDLENTIQSLRELSFKDQIHQSQINESRRRIKQTEYSEIDGRSLSSSEYDAKLKQSLQ 269

QY 184 SL-----REERFLTPLAEEVRRLL-----AHNQSLTAABELTKTIR---DSLKD 225

Db 270 DVRAQYEQMQINDEIQLS--IEDKIORLQEAARTSNSTHKSIEELRSTRVRIDALNA 327

QY 226 EIGQLSKSLTSLQIA-LQKSSD 250

Db 328 NINELEQANADLNARIDLERQLDND 353

RESULT 9

AP04_RAT STANDARD; PRT; 391 AA.

AC P02651;

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-MAR-1989 (Rel. 10, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

AP04_RAT Apolipoprotein A-IV precursor (Apo-AIV).

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=86196059; PubMed=3009456;

RA Boguski M.S., Birkenmeier E.H., Elshourbagy N.A., Taylor J.M.,

RA "Evolution of the apolipoproteins. Structure of the rat apo-A-IV gene

RT and its relationship to the human genes for apo-A-I, C-III, and E";

RL J. Biol. Chem. 261:6398-6407(1986).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=84298074; PubMed=6591177;

RA Boguski M.S., Elshourbagy N.A., Taylor J.M., Gordon J.I.,

RT "rat apolipoprotein A-IV contains 13 tandem repetitions of a 22-amino

RT acid segment with amphipathic helical potential";

RL Proc. Natl. Acad. Sci. U.S.A. 81:5021-5025(1984).

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=87008540; PubMed=3020028;

RA Haddad I.A., Ordovas J.M., Fitzpatrick T., Karathanasis S.K.,

RT "Linkage, evolution, and expression of the rat apolipoprotein A-I, C-

RT III, and A-IV genes";

RL J. Biol. Chem. 261:13268-13277(1986).

CC -1- FUNCTION: MAY HAVE A ROLE IN CHYLOMICRONS AND VLDL SECRETION AND

CC CATABOLISM. REQUIRED FOR EFFICIENT ACTIVATION OF LIPOPROTEIN

CC LIPASE BY APOC-II; POTENT ACTIVATOR OF LCAT. APOA-IV IS A MAJOR

CC COMPONENT OF HDL AND CHYLOMICRONS.

CC -1- SUBCELLULAR LOCATION: Extracellular.

CC -1- TISSUE SPECIFICITY: SECRETED IN PLASMA.

CC -1- DOMAIN: NINE OF THE THIRTEEN 22-AMINO ACID TANDEM REPEATS (EACH

CC 22-MER IS ACTUALLY A TANDEM ARRAY OF TWO A AND B, RELATED 11-

CC MERS) OCCURRING IN THIS SEQUENCE ARE PREDICTED TO BE HIGHLY ALPHA-

CC HELICAL, AND MANY OF THESE HELICES ARE AMPHIPATHIC. THEY MAY

CC THEREFORE SERVE AS LIPID-BINDING DOMAINS WITH LECITHIN:CHOLESTEROL

CC ACYLTRANSFERASE (LCAT) ACTIVATING ABILITIES.

CC -1- SIMILARITY: BELONGS TO THE APOA4 / APOA4 / APOE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; M00002; AAA85909.1; -

CC DR EMBL; J02588; AAA40747.1; -

CC DR EMBL; M13508; AAA40748.1; -

CC DR PIR; A03095; LPRTA4.

CC DR PIR; A25214; A25214.

PIR: C24700; C24700.

DR HSSP; P02671; IFZB.

DR InterPro; IPR000074; Apolipoprotein.

DR Pfam; PF01442; Apolipoprotein; 1.

KW Plasma; Lipid transport; HDL; Chylomicron; Repeat; Signal.

FT SIGNAL 1 20

FT CHAIN 21 391

FT DOMAIN 33 330

FT REPEAT 33 54

FT REPEAT 60 81

FT REPEAT 82 103

FT REPEAT 115 136

FT REPEAT 137 158

FT REPEAT 159 180

FT REPEAT 181 202

FT REPEAT 203 224

FT REPEAT 225 246

FT REPEAT 247 268

FT REPEAT 269 286

FT REPEAT 287 308

FT REPEAT 309 330

FT REPEAT 374 385

FT DOMAIN 253 253

FT VARIANT Q -> H.

SQ SEQUENCE 391 AA; 44456 MW; 24095004A809201D CRC64;

Query Match 8.4%; Score 112; DB 1; Length 391;

Best Local Similarity 21.4%; Pred. No. 1.8; Indels 74; Gaps 13;

Matches 59; Conservative 56; Mismatches 56;

QY 59 FLIAPQITVLLAL-----FTSLAGNA-----L 82

Db 2 FLKAVLTVALVATGTAQAEVTSQVAVNWDYFTQLSNNAKEAVEQLQKTDVTQQLNTL 61

QY 83 YLOKTANLHLY-QDLQREVGLSKEINFLSV---LQEFHLHLSKEFATTSKDLSA----- 133

Db 62 FQDKLGNLTAYDQLN-----KLVPFAVQLSGHLTKETKRVREIEIQELEDLRANMMPH 116

QY 134 ---VSQDFYSCLOQFRNYKGFESILDEYKNS-TEEMRKL---FSQETIADLKGSVASLR 186

Db 117 ANKVSQMGFQVQKLEHLRYATDLQAIQAQTDQMKRQLTPYIQRMQTTIQDNVENLQ 176

QY 187 BEIRFLTPLAEEVR-RLAHNQE---SLTAAIEELK-TIRSDRLDEICOLSLSKLTLSQ 240

Db 177 SS---MVFPANELKEKFNQNMGLKQLTPRANELKATIDQNLLEDLRLRLAPLAGVQEK 233

QY 241 IALQ-----RKESDLCSQIRETSLSPKRSASP 268

Db 234 LNHQMEGLAFQMKNAEELQTKVSTNIDQLQKNLAP 269

RESULT 10

VP4_ROTUM STANDARD; PRT; 775 AA.

ID VP4_ROTUM

AC P11197;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Outer capsid protein VP4 (Hemagglutinin) (Outer layer protein VP4)

DE [Contains: Outer capsid proteins VP5 and VP8].

GN S4.

OS Human rotavirus (serotype 1 / strain M37).

OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.

OC NCBI_TaxID=10954;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=88275070; PubMed=2839714;

RA Gorziglia M., Green K., Nishikawa K., Taniguchi K., Jones R.,

RA Kapikian A.Z., Chanock R.M.,

RT "Sequence of the fourth gene of human rotaviruses recovered from

RT asymptomatic or symptomatic infections.";

RL J. Virol. 62:2978-2984(1988).

RN [2]

RP SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC COMPLEX.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC ER AND THE GOLGI COMPLEX.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC -!- SIMILARITY: BELONGS TO THE VDP/USOL/YBL047C FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; X54378; CAA38253.1; -
CC ENBL; L03188; AAB00143.1; -
CC ENBL; U53668; AAB66659.1; -
CC PIR; A38455; A38455.
CC HSP; P80220; IDIP.
CC SGD; S0002216; USOL.
CC InterPro; IPR002017; Spectrin.
CC TransPort; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
CC FT DOMAIN 1 724 GLOBULAR HEAD.
CC FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
CC FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
CC FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
CC FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
CC FT CONFLICT 847 847 G -> E (IN REF. 2).
CC FT CONFLICT 924 924 E -> K (IN REF. 2).
CC FT CONFLICT 1253 1253 V -> I (IN REF. 2).
CC FT CONFLICT 1319 1319 I -> V (IN REF. 2).
CC FT CONFLICT 1461 1461 N -> S (IN REF. 2).
CC FT CONFLICT 1581 1581 G -> S (IN REF. 2).
CC FT CONFLICT 1600 1600 I -> V (IN REF. 2).
CC FT CONFLICT 1661 1661 R -> S (IN REF. 2).
CC FT CONFLICT 1772 1772 D -> DEEDDE (IN REF. 2).
CC SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;
Query Match 8.48; Score 111.5; DB 1; Length 1790;
Best Local Similarity 22.38; Pred. No. 12;
Matches 54; Conservative 42; Mismatches 81; Indels 65; Gaps 9;
Oy 74 ITSAGNALYLQKTAHLHYDLOREVGSL-----KEINFMVLVLQKEFLHLSKE-- 123
Db 964 LKSLANN-----YKDMAENESLIKAVEESKNSSIQLSNQLNKIDMSQKE 1011
Oy 124 -----FATTSKDLNSVSDFFYSCLOGFDNYKGFESLL-----D 157
Db 1012 NFQIERGSTEKNIEOLKLTISDLEQTEIKESKSDSSKDEYESQISLKEKLETATTAND 1071
Oy 158 EYKNSTEEMKRLFSOEIIADLKGVSASLREERIFLTPLAEE-VRRLAHNQESL----- 209
Db 1072 ENVNKISELTKT-RELEAEAL-AAYKNLKNELETSETSEKALKEVKEEHLKEEKIQL 1129
Oy 210 -----TAAIEELTKIR---DSLREDTGQLSOLSKTLTSQIALQKESDLCQISETLSSP 262
Db 1130 EKEATETKQOLNSLRANLESLEKEHDLAAQLKYEEQIANKEROYNEISQINDEIITST 1189
Oy 263 RK 264
Db 1190 QQ 1191
RESULT 13

GFAP_HUMAN STANDARD; PRT; 432 AA.
ID GFAP_HUMAN
AC P14136;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glial fibrillary acidic protein, astrocyte (GFAP).
GN GFAP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89296968; PubMed=2740350;
RA Reeves S.A., Helman L.J., Allison A., Israel M.A.;
RT "Molecular cloning and primary structure of human glial fibrillary
RT acidic protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:5178-5182(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90294716; PubMed=2163003;
RA Brenner M., Lampel K., Nakatani Y., Mill J., Banner C., Mearow K.,
RA Dohadwala M., Lipsky R., Freese E.;
RT "Characterization of human cDNA and genomic clones for glial
RT fibrillary acidic protein.";
RL Brain Res. Mol. Brain Res. 7:277-286(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92343304; PubMed=1636374;
RA Kumanishi T., Usui H., Ichikawa T., Nishiyama A., Katagiri T.,
RA Abe S., Yoshida Y., Washiyama K., Kuwano R., Sakimura K.;
RT "Human glial fibrillary acidic protein (GFAP): molecular cloning of
RT the complete cDNA sequence and chromosomal localization (chromosome
RT 17) of the GFAP gene.";
RL Acta Neuropathol. 83:569-578(1992).
RN [4]
RP SEQUENCE OF 352-417 FROM N.A.
RX MEDLINE=89386721; PubMed=2780570;
RA Duguid J.R., Bohmont C.W., Liu N.G., Tourtellotte W.W.;
RT "Changes in brain gene expression shared by scrapie and Alzheimer
RT disease.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:7260-7264(1989).
CC -!- FUNCTION: GFAP, A CLASS-III INTERMEDIATE FILAMENT, IS A CELL-
CC SPECIFIC MARKER THAT, DURING THE DEVELOPMENT OF THE CENTRAL
CC NERVOUS SYSTEM, DISTINGUISHES ASTROCYTES FROM OTHER GLIAL CELLS.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; J04569; AAA52528.1; -
CC ENBL; S40719; AAB22581.1; -
CC ENBL; M26638; AAA52529.1; -
CC PIR; A32936; A32936.
CC PIR; A60052; A60052.
CC HSC-2DPAGE; P14136; HUMAN.
CC MIN; 137780; -
CC InterPro; IPR001664; IF.
CC Pfam; PF00038; Filament; 1.
CC PROSITE; PS00226; IF; 1.
CC Intermediate filament; Coiled coil.
CC DOMAIN 1 72 HEAD.
CC DOMAIN 73 377 ROD.
CC DOMAIN 378 432 TAIL.
CC DOMAIN 73 104 COIL 1A.
CC DOMAIN 105 115 LINKER 1.
CC DOMAIN 116 214 COIL 1B.
FT

```

FT DOMAIN 215 230 LINKER 12.
FT DOMAIN 231 252 COIL 2A.
FT DOMAIN 253 256 LINKER 2.
FT DOMAIN 257 377 COIL 2B.
SQ SEQUENCE 432 AA; 49880 MW; E6C3B3454C3F1250 CRC64;

Query Match
Best Local Similarity 8.3%; Score 111; DB 1; Length 432;
Matches 60; Conservative 34; Mismatches 86; Indels 66; Gaps 11;

QY 82 LYLOQTANHLXQ-----DLOREVGSL-KEINFMLSVLQKFLHLSKEFA--TT 127
DQ 161 LRLAENNLAAVROADEATLRLDKIESLEEFRLKRIHEEYRELOQLARQOV 220
QY 128 SKDLSAVSDYFCSCLOGRDNYKGFES-----LLDEYKN 161
DQ 221 HVLDVAPKPDLTALKKEIRTOYEMASSNMHEAEWYRSKFADLTDAARNALLRQAKH 280
QY 162 STEEMRKLFSOEITAD---LKGVSASLREEIRFLTPLAEEVRLAHNOESLTAIEELKT 218
DQ 281 EANDYRROL-QSLTCDLSLGTNESLERQMR--EQEERHVREAASQOELARLEEBGQS 337
QY 219 IRDSLRL---DEIGQLSLSKTLTSQLQK-----ESSDLCQSIQRETLS 261
DQ 338 LKDEMARHLQEQDLLNFKYALDIEIATYRKLEGEENRITIPVQTFSNL--QIRET-SL 394
QY 262 PRKSAS 267
DQ 395 DTKSVS 400

```

```

RESULT 14
MYS3_SCHPO
ID MYS3_SCHPO STANDARD; PRT; 2104 AA.
AC O14157; 042730;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin type II heavy chain 2.
GN MYO3 OR MYP2 OR SPAC4A8.05C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=98119390; PubMed=9459302;
Moteji F., Nakano K., Kitayama C., Yamamoto M., Mabuchi I.;
"Identification of Myo3, a second type-II myosin heavy chain in the
fission yeast Schizosaccharomyces pombe.";
FEBS Lett. 420:161-166(1997).
[2]
SEQUENCE FROM N.A.
RX MEDLINE=98062346; PubMed=9398685;
Bezanilla M., Forsburg S.L., Pollard T.D.;
"Identification of a second myosin-II in Schizosaccharomyces pombe:
Myo2p is conditionally required for cytokinesis.";
Mol. Biol. Cell 8:2693-2705(1997).
[3]
SEQUENCE FROM N.A.
RX STRAIN=972;
Skellton J., Churcher C.M., Barrell B.G., Rejandream M.A., Wood V.;
Submitted (AUG-1997) to the EMBL/Genbank/DBJ databases.
-1- FUNCTION: STABILIZES THE F-ACTIN CABLES FORMING THE F-ACTIN RING
THAT SURROUNDS THE NUCLEUS DURING INTERPHASE. MAY WORK IN
CONJUNCTION WITH MYO2.
-1- SUBUNIT: BINDS TO CDC4 AND RLC1.
-1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
[2]

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB007633; BAA24579.1; -
DR EMBL; AF029788; AAC04615.1; -
DR EMBL; Z98762; CAB11475.1; -
DR HSSP; P08799; 1MND.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00242; MYSC; 1.
KW Myosin; Coiled coil; ATP-binding; Actin-binding; Alkylation.
FT DOMAIN 1 828 MYOSIN HEAD-LIKE.
FT DOMAIN 829 2104 COILED COIL (POTENTIAL).
FT DOMAIN 646 660 ACTIN-BINDING (POTENTIAL).
FT NP_BIND 182 189 ATP (POTENTIAL).
FT MOD_RES 686 686 ALKYLATION (BY SIMILARITY).
FT CONFLICT 1193 1193 D -> G (IN REF. 2).
FT CONFLICT 1304 1304 E -> G (IN REF. 2).
FT CONFLICT 1344 1344 E -> K (IN REF. 2).
FT CONFLICT 1420 1420 G -> D (IN REF. 2).
SQ SEQUENCE 2104 AA; 242570 MW; 3A0548954028D258 CRC64;

Query Match
Best Local Similarity 8.3%; Score 110.5; DB 1; Length 2104;
Matches 41; Conservative 48; Mismatches 72; Indels 27; Gaps 6;

QY 93 YDLOREVGSLKEINFMLSVLQKFLHLSKEFATTSKLSAVSQDFYSCLOGRDNYKGF 152
DQ 918 YEFLLAEKQSIIE---DLANKQTEISYLSLSLSTLEKLSIKKDEQTISSKYKELEKDY 974
QY 153 ESLDEYKNSTEEMRKLFSOEIADLKGVSASLREEIRFLTPLAEEVRLAHNOESLTA 212
DQ 975 LNMADYQHSSQHLNLSL--EKAINENLNIRELNKRLDLD-----ELLKQKSYDTK 1026
QY 213 IEELATIRDSLDEI---GQL-----SOLSKTLTSQLQKESDLCQSIQRE 257
DQ 1027 VOELREENASLKDQCRTYESQLASLVSKYSETSELNKK-EAELVIFQKEITEYRDQLHK 1085
QY 258 TLSSPRKS 265
DQ 1086 AFQNPKEPT 1093

RESULT 15
VP4_ROTHT
ID VP4_ROTHT STANDARD; PRT; 774 AA.
AC P11200;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer capsid protein VP4 (Hemagglutinin) (Outer layer protein VP4)
DE [Contains: Outer capsid proteins VP5 and VP8].
GN S4.
OS Human rotavirus (serotype 4 / strain St. Thomas 3).
OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX NCBI_TaxID=10960;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=88275070; PubMed=2839714;
Gorziglia M., Green K., Nishikawa K., Taniguchi K., Jones R.,
Rakipkian A.Z., Chanock R.M.;
"Sequence of the fourth gene of human rotaviruses recovered from
RT asymptomatic or symptomatic infections.";
RL J. Virol. 62:2978-2984(1988).
[2]

```

RP SEQUENCE OF 1-280.
RX MEDLINE=86313706; PubMed=3018754;
RA Gorziglia M., Hoshino Y., Buckler-White A., Blumentals I., Glass R.,
RA Flores J., Kapikian A.Z., Chanock R.M.;
RT "Conservation of amino acid sequence of VP8 and cleavage region of
RT 84-kDa outer capsid protein among rotaviruses recovered from
RT asymptomatic neonatal infection";
RL Proc Natl Acad Sci U.S.A. 83:7039-7043(1986).
CC -1- SUBCELLULAR LOCATION: Outer capsid.
CC -1- PTM: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE
CC OTHER PRODUCT IS VP5.
CC -1- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.

DR PIR; H28839; VPRXWT.
DR InterPro: IPR000416; Cap_VP4.
DR Pfam: PF00426; VP4; 1.
KW Coat protein; glycoprotein.
FT CHAIN 1 774 OUTER CAPSID PROTEIN VP4.
FT CHAIN 1 240 OUTER CAPSID PROTEIN VP8.
FT CHAIN 247 774 OUTER CAPSID PROTEIN VP5.
FT CARBOHYD 32 32 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 132 132 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 583 583 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 606 606 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 774 AA; 87455 MW; D397E5C6125A8FA8 CRC64;

Query Match 8.3%; Score 110; DB 1; Length 774;
Best Local Similarity 25.3%; Pred. No. 5.4; Indels 32; Caps 7;
Matches 47; Conservative 31; Mismatches 76; Indels 32; Caps 7;
QY 94 QDLOREVGSLKEINFMLSVLQKEFLHLSKEFATTSK-DLSAVSQDFYSCLOGFRDNYKGF 152
Db 491 QDLERQLGLDRE-----EFNSLSQEIATITQIDLALLPLDMFSM-----FSGI 533
QY 153 ESSLDEYKNSTEMRKLFQSQEIADLKGVSASLREIRFLPLAEVVR---LAHQESL 209
Db 534 KSTIDAAKSMATKVMKKFR-----SGLATSISETLSLNAASSVSRSSIRSNISSI 587
QY 210 ---TAAIEELKTIRDSLRDEIGLSQSLKTLTSQIALQKSS--DLCSQIRETLSSPRK 264
Db 588 SEWTDVSEQIAGSSDSVRNISTQTSASRRLRLREITTTQEGMNFIDISAAVLKTKIDRST 647
QY 265 SASPST 270
Db 648 HIRPDT 653

Search completed: August 13, 2002, 09:30:44
Job time: 380 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 09:30:16 ; Search time 48.57 Seconds
(without alignments)
972.362 Million cell updates/sec

Title: US-09-673-763-14
Perfect score: 1333
Sequence: 1 MTTPTLVIPSPAPSYA.....QIRETLSPKSPSTKSS 273

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1333	100.0	273	2	O69196
2	1328	99.6	273	2	Q9A5A5
3	1318	98.9	273	2	Q9A5A5
4	1315	98.6	273	2	Q9A5A5
5	1311	98.3	273	16	O84121
6	1310	98.3	273	2	Q9A5A5
7	1308	98.1	273	2	Q9A5A5
8	1307	98.0	273	2	Q9A5A5
9	1306	98.0	273	2	Q9A5A5
10	1306	98.0	273	2	Q9A5A5
11	1305	97.9	273	2	Q9A5A5
12	1305	97.9	273	2	Q9A5A5
13	1302	97.7	273	2	Q9A5A5
14	1293	97.0	273	2	Q9A5A5
15	1292	96.9	273	2	Q9A5A5
16	1209.5	90.7	256	2	Q9F7L1

17	962	72.2	197	2	Q9F7T6	Q9F7T6 chlamydia t
18	756.5	56.8	158	2	Q9F7L0	Q9F7L0 chlamydia t
19	714	53.6	276	16	Q9PKR8	Q9PKR8 chlamydia m
20	612	45.9	174	2	Q9F7L2	Q9F7L2 chlamydia t
21	582	43.7	121	2	Q9AM93	Q9AM93 chlamydia t
22	564	42.3	174	2	Q9F7K8	Q9F7K8 chlamydia t
23	503	37.7	109	2	Q9F7M7	Q9F7M7 chlamydia t
24	162.5	12.2	355	2	Q46210	Q46210 chlamydia t
25	154.5	11.6	390	16	Q92828	Q92828 chlamydia p
26	128.5	9.6	472	13	Q07962	Q07962 carassius a
27	126	9.5	1179	17	O59462	O59462 pyrococcus
28	124	9.3	263	13	Q98TG6	Q98TG6 anguilla ja
29	124	9.3	347	3	Q02428	Q02428 saccharomyc
30	123.5	9.3	471	13	P87360	P87360 brachydanio
31	122.5	9.2	3259	4	Q14789	Q14789 homo sapien
32	122	9.2	24	2	Q9F7T8	Q9F7T8 chlamydia t
33	122	9.2	775	12	Q86200	Q86200 human rotav
34	122	9.2	1931	5	Q9NCF9	Q9NCF9 drosophila
35	120.5	9.0	511	13	Q90ZF7	Q90ZF7 rana catesb
36	120.5	9.0	718	4	Q9BV13	Q9BV13 homo sapien
37	120.5	9.0	1286	5	Q21025	Q21025 caenorhabdi
38	120	9.0	644	16	Q9KMZ0	Q9KMZ0 vibrio chol
39	120	9.0	1931	5	Q9VAK9	Q9VAK9 drosophila
40	119.5	9.0	406	2	Q9L514	Q9L514 salmonella
41	119.5	9.0	1743	5	O96063	O96063 dugesia jap
42	119	8.9	897	13	O13098	O13098 xenopus lae
43	117.5	8.8	775	12	O90314	O90314 human rotav
44	117.5	8.8	1475	5	O76447	O76447 caenorhabdi
45	117	8.8	1101	4	Q96SB9	Q96SB9 homo sapien

ALIGNMENTS

RESULT 1

O69196 PRELIMINARY; PRT; 273 AA.
ID O69196;
AC O69196;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE INCLUSION MEMBRANE PROTEIN.
GN INCA.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LGV-434;
RX MEDLINE=99043938; PubMed=9826388;
RA Bannantine J.P., Stamm W.E., Suchland R.J., Rockey D.D.;
RT "Chlamydia trachomatis Inca is localized to the inclusion membrane and
is recognized by antisera from infected humans and primates.";
RL Infect. Immun. 66:6017-6021(1998).
DR EMBL; AF0067958; AAC82641.1;
SQ SEQUENCE 273 AA; 30272 MW; E64B846A5D6824EC CRC64;

Query Match 100.0%; Score 1333; DB 2; Length 273;
Best Local Similarity 100.0%; Pred. No. 1.6e-81;
Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MTTPTLVIPSPAPSYANRVPPQPSLMKIKKIAIASLILGTIGTIGLALGHVGL 60	
Db	1	MTTPTLVIPSPAPSYANRVPPQPSLMKIKKIAIASLILGTIGTIGLALGHVGL 60	
Qy	61	IAPQITIVLLALFTTSLAGNALYLQKTANLHLYQDLQREVGSLSKEINFMLSVLQKEFLH 120	
Db	61	IAPQITIVLLALFTTSLAGNALYLQKTANLHLYQDLQREVGSLSKEINFMLSVLQKEFLH 120	
Qy	121	SKEFATTSKDLSSVQDFYSCLOGFRDNYKGFESLLDEYNKSTEEMRKLFQSQIIADLKG 180	
Db	121	SKEFATTSKDLSSVQDFYSCLOGFRDNYKGFESLLDEYNKSTEEMRKLFQSQIIADLKG 180	

RP SEQUENCE FROM N.A.
RC STRAIN-TW-5;
RX MEDLINE=21295121; PubMed=11402010;
RA Pannekoek Y., van Der Ende A., Eijk P.P., van Marle J., de Witte M.A.,
RA Ossewaarde J.M., van Den Brule A.J.C., Morre S.A., Dankert J.;
RT "Normal Inca expression and fusogenicity of inclusions in Chlamydia
trachomatis isolates with the inca 147T mutation.";
RL Infect. Immun. 69:4654-4656(2001).
DR EMBL: AF326993; AAG61090.1; -.
SQ SEQUENCE 273 AA; 30318 MW; 22729569405E422B CRC64;

Query Match 98.9%; Score 1318; DB 2; Length 273;
Best Local Similarity 98.9%; Pred. No. 1.6e-80;
Matches 270; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTPTTLIVIPPPAPPSANRVQPQLMDKIKKIAAIAASLLIGTIGFLALHGVGL 60
Db 1 MTPTTLIVIPPPAPPSANRVQPQLMDKIKKIAAIAASLLIGTIGFLALHGVGL 60

QY 61 IAPQITIVLLALFITSAGNALYLOKTANHLHYQDLQREVGLKEINFMVLSVLOKEFLHL 120
Db 61 IAPQITIVLLALFITSAGNALYLOKTANHLHYQDLQREVGLKEINFMVLSVLOKEFLHL 120

QY 121 SKEFATSKDLSAVSQDFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFQSEIIADLKG 180
Db 121 SKEFATSKDLSAVSQDFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFQSEIIADLKS 180

QY 181 SVASLREEIRFLTPLAEEVRLAHNOESLTAATEELKTIRDSLRDEIGQLSKLTTSQ 240
Db 181 SVASLREEIRFLTPLAEEVRLAHNOESLTAATEELKTIRDSLRDEIGQLSKLTTSQ 240

QY 241 IALQKESDLCQIRETLSSPRKSASPSTKSS 273
Db 241 IALQKESDLCQIRETLSSPRKSASPSTKSS 273

RESULT 4
Q9AMA4 PRELIMINARY; PRT; 273 AA.
ID Q9AMA4
AC Q9AMA4; 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE INCLUSION MEMBRANE PROTEIN A.
GN INCA.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-404-L;
RX MEDLINE=21295121; PubMed=11402010;
RA Pannekoek Y., van Der Ende A., Eijk P.P., van Marle J., de Witte M.A.,
RA Ossewaarde J.M., van Den Brule A.J.C., Morre S.A., Dankert J.;
RT "Normal Inca expression and fusogenicity of inclusions in Chlamydia
trachomatis isolates with the inca 147T mutation.";
RL Infect. Immun. 69:4654-4656(2001).
DR EMBL: AF327010; AAG61107.1; -.
SQ SEQUENCE 273 AA; 30346 MW; 7E9F9569405E5469 CRC64;

QY 181 SVASLREEIRFLTPLAEEVRLAHNOESLTAATEELKTIRDSLRDEIGQLSKLTTSQ 240
Db 181 SVASLREEIRFLTPLAEEVRLAHNOESLTAATEELKTIRDSLRDEIGQLSKLTTSQ 240

QY 241 IALQKESDLCQIRETLSSPRKSASPSTKSS 273
Db 241 IALQKESDLCQIRETLSSPRKSASPSTKSS 273

RESULT 2
Q9AMA5 PRELIMINARY; PRT; 273 AA.
ID Q9AMA5
AC Q9AMA5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE INCLUSION MEMBRANE PROTEIN A.
GN INCA.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-434/BU;
RX MEDLINE=21295121; PubMed=11402010;
RA Pannekoek Y., van Der Ende A., Eijk P.P., van Marle J., de Witte M.A.,
RA Ossewaarde J.M., van Den Brule A.J.C., Morre S.A., Dankert J.;
RT "Normal Inca expression and fusogenicity of inclusions in Chlamydia
trachomatis isolates with the inca 147T mutation.";
RL Infect. Immun. 69:4654-4656(2001).
DR EMBL: AF327009; AAG61106.1; -.
SQ SEQUENCE 273 AA; 30259 MW; E0B3496A529B14EC CRC64;

Query Match 99.6%; Score 1328; DB 2; Length 273;
Best Local Similarity 99.6%; Pred. No. 3.4e-81;
Matches 272; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTPTTLIVIPPPAPPSANRVQPQLMDKIKKIAAIAASLLIGTIGFLALHGVGL 60
Db 1 MTPTTLIVIPPPAPPSANRVQPQLMDKIKKIAAIAASLLIGTIGFLALHGVGL 60

QY 61 IAPQITIVLLALFITSAGNALYLOKTANHLHYQDLQREVGLKEINFMVLSVLOKEFLHL 120
Db 61 IAPQITIVLLALFITSAGNALYLOKTANHLHYQDLQREVGLKEINFMVLSVLOKEFLHL 120

QY 121 SKEFATSKDLSAVSQDFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFQSEIIADLKG 180
Db 121 SKEFATSKDLSAVSQDFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFQSEIIADLKG 180

QY 181 SVASLREEIRFLTPLAEEVRLAHNOESLTAATEELKTIRDSLRDEIGQLSKLTTSQ 240
Db 181 SVASLREEIRFLTPLAEEVRLAHNOESLTAATEELKTIRDSLRDEIGQLSKLTTSQ 240

QY 241 IALQKESDLCQIRETLSSPRKSASPSTKSS 273
Db 241 IALQKESDLCQIRETLSSPRKSASPSTKSS 273

RESULT 3
Q9AMB2 PRELIMINARY; PRT; 273 AA.
ID Q9AMB2
AC Q9AMB2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE INCLUSION MEMBRANE PROTEIN A.
GN INCA.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]

SQ SEQUENCE 273 AA; 30357 MW; 61FA8F2085701F58 CRC64;

Query Match 98.0%; Score 1307; DB 2; Length 273;
Best Local Similarity 97.8%; Pred. No. 8.5e-80;
Matches 267; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 MTTPTLIVIPPPAPPSANRVPQPSLMDKIKKIAAIAASLILIGTIGFLALLGHLVGL 60
Db 1 MTTPTLIVIPPPAPPSANRVPQPSLMDKIKKIAAIAASLILIGTIGFLALLGHLVGL 60
QY 61 IAPQITIVLLALFITSLAGNALYLQKTANLHLYQDLQREVGSKEINFMVLSVLOKEFLHL 120
Db 61 IAPQITIVLLALFITSLAGNALYLQKTANLHLYQDLQREVGSKEINFMVLSVLOKEFLHL 120
QY 121 SKEFATTSKDLASVQDFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFSOEIIADLKG 180
Db 121 SKEFATTSKDLASVQDFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFSOEIIADLKG 180
QY 181 SVASLREEIRFTPLAEVRRRLAHNQESLTAIEELKTIIRDSLRDEIGOLSQSLKTLTSQ 240
Db 181 SVASLREEIRFTPLAEVRRRLAHNQESLTVVIEELKTIIRDSLRDEIGOLSQSLKTLTSQ 240
QY 241 IALQKESDLCQIRETLSSPRKSASPSTKSS 273
Db 241 IALQKESDLCQIRETLSSPRKSASPSTKSS 273

RESULT 9
Q9F7K9 PRELIMINARY; PRT; 273 AA.
ID Q9F7K9 AC Q9F7K9
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE INCLUSION MEMBRANE PROTEIN A.
GN INCA.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT5942;
RA Viratyosin W., Rockey D.D., Suchland R.J., Stamm W.E.;
RT "Diverse Mutations in Inca Amplified from Clinical Chlamydia
trachomatis Isolates that Occupy Nonfusogenic Inclusions.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=7A, DK-20, MRC-301, 4A, AND 12A;
RA Pannekoek Y., van der Ende A., Eijk P.P., van Marle J., de Witte M.A.,
RA Ossewaarde J.M., Morre S.A., van den Brule A.J.C., Dankert J.;
RT "Normal Inca expression and fusogenicity of inclusions in Chlamydia
trachomatis isolates with the Inca 147T mutation.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF279354; AAG31469.1; -
DR EMBL; AF327016; AAG61113.1; -
DR EMBL; AF327000; AAG61097.1; -
DR EMBL; AF327001; AAG61098.1; -
DR EMBL; AF327011; AAG61108.1; -
DR EMBL; AF327013; AAG61110.1; -
SQ SEQUENCE 273 AA; 30315 MW; 2986A622F1B3BCE6 CRC64;

Query Match 98.0%; Score 1306; DB 2; Length 273;
Best Local Similarity 97.8%; Pred. No. 9.9e-80;
Matches 267; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 MTTPTLIVIPPPAPPSANRVPQPSLMDKIKKIAAIAASLILIGTIGFLALLGHLVGL 60
Db 1 MTTPTLIVIPPPAPPSANRVPQPSLMDKIKKIAAIAASLILIGTIGFLALLGHLVGL 60
QY 61 IAPQITIVLLALFITSLAGNALYLQKTANLHLYQDLQREVGSKEINFMVLSVLOKEFLHL 120

Db 241 IALQKESDLCQIRETLSSPRKSASPSTKSS 273

RESULT 7
Q9AM94 PRELIMINARY; PRT; 273 AA.
ID Q9AM94 AC Q9AM94
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE INCLUSION MEMBRANE PROTEIN A.
GN INCA.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT9227;
RA Viratyosin W., Rockey D.D., Suchland R.J., Stamm W.E.;
RT "Diverse Mutations in Inca Amplified from Clinical Chlamydia
trachomatis Isolates that Occupy Nonfusogenic Inclusions.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF327326; AAK11020.1; -
SQ SEQUENCE 273 AA; 30354 MW; 6945AE6A0D5092C3 CRC64;

Query Match 98.1%; Score 1308; DB 2; Length 273;
Best Local Similarity 97.8%; Pred. No. 7.3e-80;
Matches 267; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 1 MTTPTLIVIPPPAPPSANRVPQPSLMDKIKKIAAIAASLILIGTIGFLALLGHLVGL 60
Db 1 MTTPTLIVIPPPAPPSANRVPQPSLMDKIKKIAAIAASLILIGTIGFLALLGHLVGL 60
QY 61 IAPQITIVLLALFITSLAGNALYLQKTANLHLYQDLQREVGSKEINFMVLSVLOKEFLHL 120
Db 61 IAPQITIVLLALFITSLAGNALYLQKTANLHLYQDLQREVGSKEINFMVLSVLOKEFLHL 120
QY 121 SKEFATTSKDLASVQDFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFSOEIIADLKG 180
Db 121 SKEFATTSKDLASVQDFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFSOEIIADLKG 180
QY 181 SVASLREEIRFTPLAEVRRRLAHNQESLTAIEELKTIIRDSLRDEIGOLSQSLKTLTSQ 240
Db 181 SVASLREEIRFTPLAEVRRRLAHNQESLTVVIEELKTIIRDSLRDEIGOLSQSLKTLTSQ 240
QY 241 IALQKESDLCQIRETLSSPRKSASPSTKSS 273
Db 241 IALQKESDLCQIRETLSSPRKSASPSTKSS 273

RESULT 8
Q9AMA9 PRELIMINARY; PRT; 273 AA.
ID Q9AMA9 AC Q9AMA9
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE INCLUSION MEMBRANE PROTEIN A.
GN INCA.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UW-4;
RX MEDLINE=21295121; PubMed=11402010;
RA Pannekoek Y., van der Ende A., Eijk P.P., van Marle J., de Witte M.A.,
RA Ossewaarde J.M., van den Brule A.J.C., Morre S.A., Dankert J.;
RT "Normal Inca expression and fusogenicity of inclusions in Chlamydia
trachomatis isolates with the Inca 147T mutation.";
RL Infect. Immun. 69:4654-4656(2001).
DR EMBL; AF327003; AAG61100.1; -


```

Db 61 IAPQITIVLLALFIISLGNALYLQKTNLHLYQDLQREVGLKEINFMLSVLQKEFLHL 120
QY 121 SKEFATTSKDLSSAVSQDFYSCLOGFRDNYKGFESLLDEYKNSTEEMKRLFSQBIADLKG 180
Db 121 SKEFATTSKDLSSAVSQDFYSCLOGFRDNYKGFESLLDEYKNSTEEMKRLFSQBIADLKG 180
QY 181 SVASLREEIRFLTPLAEEVRLAHNOESLTAATIEELKTIIRDSLRDEIGQLSLSKLTLSQ 240
Db 181 SVASLREEIRFLTPLAEEVRLAHNOESLTAATIEELKTIIRDSLRDEIGQLSLSKLTLSQ 240
QY 241 IALQKESSDLCQIRETLSSPRKSASPSTKSS 273
Db 241 IALQKESSDLCQIRETLSSPRKSASPSTKSS 273

RESULT 10
Q9AMA7
ID Q9AMA7 PRELIMINARY; PRT; 273 AA.
AC Q9AMA7;
QY 01-JUN-2001 (TrEMBLrel. 17, Created)
QY 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
QY 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE INCLUSION MEMBRANE PROTEIN A.
GN INCA.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UW-31;
RX MEDLINE=21295121; PubMed=11402010;
RA Pannekoek Y., van Der Ende A., Eljk P.P., van Marle J., de Witte M.A.,
RA Ossewaarde J.M., van Den Brule A.J.C., Morre S.A., Dankert J.;
RT "Normal Inca expression and fusogenicity of inclusions in Chlamydia
RT trachomatis isolates with the inca I47T mutation.";
RL Infect. Immun. 69:4654-4656(2001).
DR EMBL; AF327007; AAG61104.1; -.
SQ SEQUENCE 273 AA; 30337 MW; 6945AE7E8B8824F3 CRC64;

```

```

Query Match 98.0%; Score 1306; DB 2; Length 273;
Best Local Similarity 97.8%; Pred. No. 9.9e-80;
Matches 267; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 MTTPTLIVTPPPAPSPYSANRVPQPSLMKDKIKKIAAIAASLILIGTIGFLALLGHLVGL 60
Db 1 MTTPTLIVTPPPAPSPYSANRVPQPSLMKDKIKKIAAIAASLILIGTIGFLALLGHLVGL 60
QY 61 IAPQITIVLLALFIISLGNALYLQKTNLHLYQDLQREVGLKEINFMLSVLQKEFLHL 120
Db 61 IAPQITIVLLALFIISLGNALYLQKTNLHLYQDLQREVGLKEINFMLSVLQKEFLHL 120
QY 121 SKEFATTSKDLSSAVSQDFYSCLOGFRDNYKGFESLLDEYKNSTEEMKRLFSQBIADLKG 180
Db 121 SKEFATTSKDLSSAVSQDFYSCLOGFRDNYKGFESLLDEYKNSTEEMKRLFSQBIADLKG 180
QY 181 SVASLREEIRFLTPLAEEVRLAHNOESLTAATIEELKTIIRDSLRDEIGQLSLSKLTLSQ 240
Db 181 SVASLREEIRFLTPLAEEVRLAHNOESLTAATIEELKTIIRDSLRDEIGQLSLSKLTLSQ 240
QY 241 IALQKESSDLCQIRETLSSPRKSASPSTKSS 273
Db 241 IALQKESSDLCQIRETLSSPRKSASPSTKSS 273

RESULT 11
Q99Q56
ID Q99Q56 PRELIMINARY; PRT; 273 AA.
AC Q99Q56;
QY 01-JUN-2001 (TrEMBLrel. 17, Created)
QY 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

```

```

DE INCLUSION MEMBRANE PROTEIN A.
GN INCA.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WT9329, AND MT9309;
RA Viratyosin W., Rocky D.D., Suchland R.J., Stamm W.E.;
RT "Diverse Mutations in Inca Amplified from Clinical Chlamydia
RT trachomatis Isolates that Occupy Nonfusogenic Inclusions.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DDJ databases.
DR EMBL; AF327331; AAK11235.1; -.
DR EMBL; AF327330; AAK11234.1; -.
SQ SEQUENCE 273 AA; 30339 MW; B445B3638B8BEBE7 CRC64;

Query Match 97.9%; Score 1305; DB 2; Length 273;
Best Local Similarity 97.8%; Pred. No. 1.2e-79;
Matches 267; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 MTTPTLIVTPPPAPSPYSANRVPQPSLMKDKIKKIAAIAASLILIGTIGFLALLGHLVGL 60
Db 1 MTTPTLIVTPPPAPSPYSANRVPQPSLMKDKIKKIAAIAASLILIGTIGFLALLGHLVGL 60
QY 61 IAPQITIVLLALFIISLGNALYLQKTNLHLYQDLQREVGLKEINFMLSVLQKEFLHL 120
Db 61 IAPQITIVLLALFIISLGNALYLQKTNLHLYQDLQREVGLKEINFMLSVLQKEFLHL 120
QY 121 SKEFATTSKDLSSAVSQDFYSCLOGFRDNYKGFESLLDEYKNSTEEMKRLFSQBIADLKG 180
Db 121 SKEFATTSKDLSSAVSQDFYSCLOGFRDNYKGFESLLDEYKNSTEEMKRLFSQBIADLKG 180
QY 181 SVASLREEIRFLTPLAEEVRLAHNOESLTAATIEELKTIIRDSLRDEIGQLSLSKLTLSQ 240
Db 181 SVASLREEIRFLTPLAEEVRLAHNOESLTAATIEELKTIIRDSLRDEIGQLSLSKLTLSQ 240
QY 241 IALQKESSDLCQIRETLSSPRKSASPSTKSS 273
Db 241 IALQKESSDLCQIRETLSSPRKSASPSTKSS 273

RESULT 12
Q9AMA8
ID Q9AMA8 PRELIMINARY; PRT; 273 AA.
AC Q9AMA8;
QY 01-JUN-2001 (TrEMBLrel. 17, Created)
QY 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
QY 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE INCLUSION MEMBRANE PROTEIN A.
GN INCA.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UW-12;
RX MEDLINE=21295121; PubMed=11402010;
RA Pannekoek Y., van Der Ende A., Eljk P.P., van Marle J., de Witte M.A.,
RA Ossewaarde J.M., van Den Brule A.J.C., Morre S.A., Dankert J.;
RT "Normal Inca expression and fusogenicity of inclusions in Chlamydia
RT trachomatis isolates with the inca I47T mutation.";
RL Infect. Immun. 69:4654-4656(2001).
DR EMBL; AF327005; AAG61102.1; -.
SQ SEQUENCE 273 AA; 30387 MW; 6945AE7E9B8AEBF3 CRC64;

Query Match 97.9%; Score 1305; DB 2; Length 273;
Best Local Similarity 97.8%; Pred. No. 1.2e-79;
Matches 267; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 MTTPTLIVTPPPAPSPYSANRVPQPSLMKDKIKKIAAIAASLILIGTIGFLALLGHLVGL 60

```

```
Db 1 MTTPTLIVPPSPAPSYSANRVPQPSLMDKIKKIAIASLILIGTIGFLLGLHVL 60
QY 61 IAPQITIVLLALFITSAGNALYLOKTANHLVQDLQREVGSKEINFMVSVLQKEFLHL 120
Db 61 IAPQITIVLLALFITSAGNALYLOKTANHLVQDLQREVGSKEINFMVSVLQKEFLHL 120
QY 121 SKEFATTSKDLASVQDFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFQSEIIADLKG 180
Db 121 SKEFATTSKDLASVQDFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFQSEIIADLKG 180
QY 121 SKEFATTSKDLASVQDFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFQSEIIADLKG 180
Db 121 SKEFATTSKDLASVQDFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFQSEIIADLKG 180
QY 181 SVASLREEIRFLTPLAEEVRRLAHNOESLTAATIEELKTIRDSLRDEIGOLSQSLKTLTSQ 240
Db 181 SVASLREEIRFLTPLAEEVRRLAHNOESLTAATIEELKTIRDSLRDEIGOLSQSLKTLTSQ 240
QY 241 IALQKESDLCQIRETLSSPRKSASPSTKSS 273
Db 241 IALQKESDLCQIRETLSSPRKSASPSTKSS 273

RESULT 13
Q9RFX7
ID Q9RFX7 PRELIMINARY; PRT; 273 AA.
AC Q9RFX7
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE INCLUSION MEMBRANE PROTEIN A.
GN INCA.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D(S) 2923;
RX MEDLINE=20072706; PubMed=10603409;
RA "Isolates of Chlamydia trachomatis that occupy nonfusogenic inclusions lack Inca, a protein localized to the inclusion membrane."
RL Infect. Immun. 68:360-367(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MT9346, MT9334, MT9291, AND MT9336;
RA Pannekoek Y., van der Ende A., Eijk P.P., van Marle J., de Witte M.A., Viratytosin W., Rockey D.D., Suchland R.J., Stamm W.E.;
RT "Diverse Mutations in Inca Amplified from Clinical Chlamydia trachomatis Isolates that Occupy Nonfusogenic Inclusions."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN=IC-CAL8;
RA Pannekoek Y., van der Ende A., Eijk P.P., van Marle J., de Witte M.A., Ossewaarde J.M., Morre S.A., van den Brule A.J.C., Dankert J.;
RT "Normal Inca expression and fusogenicity of inclusions in Chlamydia trachomatis isolates with the Inca 147T mutation."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF163773; AAD48440.1; -
DR EMBL; AF327333; AAK11237.1; -
DR EMBL; AF326996; AAG61093.1; -
DR EMBL; AF326999; AAG61096.1; -
DR EMBL; AF327327; AAK11231.1; -
DR EMBL; AF327328; AAK11232.1; -
DR EMBL; AF327332; AAK11236.1; -
SQ SEQUENCE 273 AA; 30314 MW; 9D0C862A533FDE8C CRC64;

Query Match 97.7%; Score 1302; DB 2; Length 273;
Best Local Similarity 97.4%; Pred. No. 1.8e-79;
Matches 266; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MTTPTLIVPPSPAPSYSANRVPQPSLMDKIKKIAIASLILIGTIGFLLGLHVL 60
Db 1 MTTPTLIVPPSPAPSYSANRVPQPSLMDKIKKIAIASLILIGTIGFLLGLHVL 60
QY 61 IAPQITIVLLALFITSAGNALYLOKTANHLVQDLQREVGSKEINFMVSVLQKEFLHL 120
```

```
Db 61 IAPQITIVLLALFITSAGNALYLOKTANHLVQDLQREVGSKEINFMVSVLQKEFLHL 120
QY 121 SKEFATTSKDLASVQDFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFQSEIIADLKG 180
Db 121 SKEFATTSKDLASVQDFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFQSEIIADLKG 180
QY 181 SVASLREEIRFLTPLAEEVRRLAHNOESLTAATIEELKTIRDSLRDEIGOLSQSLKTLTSQ 240
Db 181 SVASLREEIRFLTPLAEEVRRLAHNOESLTAATIEELKTIRDSLRDEIGOLSQSLKTLTSQ 240
QY 241 IALQKESDLCQIRETLSSPRKSASPSTKSS 273
Db 241 IALQKESDLCQIRETLSSPRKSASPSTKSS 273

RESULT 14
Q9AMB0
ID Q9AMB0 PRELIMINARY; PRT; 273 AA.
AC Q9AMB0;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE INCLUSION MEMBRANE PROTEIN A.
GN INCA.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT-566;
RX MEDLINE=21295121; PubMed=11402010;
RA Pannekoek Y., van der Ende A., Eijk P.P., van Marle J., de Witte M.A., Ossewaarde J.M., van den Brule A.J.C., Morre S.A., Dankert J.;
RT "Normal Inca expression and fusogenicity of inclusions in Chlamydia trachomatis isolates with the Inca I47T mutation."
RL Infect. Immun. 69:4654-4656(2001).
DR EMBL; AF326997; AAG61094.1; -
SQ SEQUENCE 273 AA; 30361 MW; 42C614AD71089CF0 CRC64;

Query Match 97.0%; Score 1293; DB 2; Length 273;
Best Local Similarity 97.4%; Pred. No. 7.3e-79;
Matches 266; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MTTPTLIVPPSPAPSYSANRVPQPSLMDKIKKIAIASLILIGTIGFLLGLHVL 60
Db 1 MTTPTLIVPPSPAPSYSANRVPQPSLMDKIKKIAIASLILIGTIGFLLGLHVL 60
QY 61 IAPQITIVLLALFITSAGNALYLOKTANHLVQDLQREVGSKEINFMVSVLQKEFLHL 120
Db 61 IAPQITIVLLALFITSAGNALYLOKTANHLVQDLQREVGSKEINFMVSVLQKEFLHL 120
QY 121 SKEFATTSKDLASVQDFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFQSEIIADLKG 180
Db 121 SKEFATTSKDLASVQDFYSCLOGFRDNYKGFESLLDEYKNSTEMRKLFQSEIIADLKG 180
QY 181 SVASLREEIRFLTPLAEEVRRLAHNOESLTAATIEELKTIRDSLRDEIGOLSQSLKTLTSQ 240
Db 181 SVASLREEIRFLTPLAEEVRRLAHNOESLTAATIEELKTIRDSLRDEIGOLSQSLKTLTSQ 240
QY 241 IALQKESDLCQIRETLSSPRKSASPSTKSS 273
Db 241 IALQKESDLCQIRETLSSPRKSASPSTKSS 273

RESULT 15
Q9AMB1
ID Q9AMB1 PRELIMINARY; PRT; 273 AA.
AC Q9AMB1;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
```

Search completed: August 13, 2002, 09:30:17
Job time: 373 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 09:22:24 ; Search time 29.79 Seconds
(without alignments)
1145.073 Million cell updates/sec

Title: US-09-673-763-8
Perfect score: 1733
Sequence: 1 MTVSTNTSPVISRASSPTF.....DLSAQPDQNSDAGEHKDS 355

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1733	100.0	355	2 S61491	inclusion membrane
2	200	11.5	390	2 A72108	hypothetical prote
3	200	11.5	390	2 B85114	hypothetical prote
4	185	10.7	276	2 H81707	inclusion membrane
5	164.5	9.5	273	2 C71553	probable inclusion
6	149	8.6	533	2 G72593	hypothetical prote
7	145.5	8.4	1732	2 T14039	protein kinase (EC
8	145	8.4	1690	2 T13030	microtubule bindin
9	143	8.3	1410	1 A57013	spindle pole body
10	141	8.1	944	2 S26710	outer membrane pro
11	140.5	8.1	403	2 H72216	methyl-accepting c
12	140	8.1	770	2 F82383	chromosome segrega
13	140	8.1	1056	1 G02157	structural maintain
14	140	8.1	1155	2 A2675	myosin II heavy ch
15	140	8.1	1165	2 A37457	hypothetical prote
16	139.5	8.0	739	2 H75001	hypothetical prote
17	139.5	8.0	746	2 T47237	cytron - mouse
18	139	8.0	1805	1 A64224	skeletal myosin -
19	138.5	8.0	772	2 T27907	paramyosin - fluke
20	138.5	8.0	1597	2 S68420	hypothetical prote
21	138.5	8.0	1957	2 A39294	myosin heavy chain
22	137	7.9	439	2 A60608	myosin heavy chain
23	136.5	7.9	1133	2 T22976	male-enhanced anti
24	136.5	7.9	1940	1 S04090	apsB protein - Eme
25	136.5	7.9	1940	1 A24922	myosin heavy chain
26	136	7.8	1325	2 T42722	interaptin - slime
27	135.5	7.8	1051	2 T18302	
28	135.5	7.8	1937	2 I38055	
29	135	7.8	1738	2 T14867	

ALIGNMENTS

RESULT 1

S61491

inclusion membrane protein A - Chlamydomophila psittaci

C:Species: Chlamydomophila psittaci, Chlamydia psittaci

C>Date: 27-Apr-1996 #sequence_revision 13-Mar-1997 #text_change 31-Mar-2000

C:Accession: S61491

R:Rockey, D.D.; Heinzen, R.A.; Hackstadt, T.

Mol. Microbiol. 15, 617-626, 1995

A:Title: Cloning and characterization of a Chlamydia psittaci gene coding for a prote

A:Reference number: S61491; MUID:95302975

A:Accession: S61491

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-355 <R>

A:Cross-references: EMBL:L35036; NID:g516598; PIDN:AAC41443.1; PID:g516599

Query Match 100.0%; Score 1733; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 4.3e-93;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTVSTNTSPVISRASSPTFGDHGKDFDNKIIPIISIEAPTSSAAAVGAKTAIEPEGKSP 60
DB 1 MTVSTNTSPVISRASSPTFGDHGKDFDNKIIPIISIEAPTSSAAAVGAKTAIEPEGKSP 60
QY 61 LLQRCYLVKIIATAIAFWGIAALVCLYLGSVISTPSSLILMLAIVLVSFVIVITARDG 120
DB 61 LLQRCYLVKIIATAIAFWGIAALVCLYLGSVISTPSSLILMLAIVLVSFVIVITARDG 120
QY 121 TPSQVVRHMKQIQOFGENTRLHTAVENKAVNVELSEQINQLKHLTRLSDFGDRLEA 180
DB 121 TPSQVVRHMKQIQOFGENTRLHTAVENKAVNVELSEQINQLKHLTRLSDFGDRLEA 180
QY 181 NTGFTALIAIDFQLSLEEFKSVGKVKETMLSPFEKLAQSLKETFSQAVQAMMSVTELR 240
DB 181 NTGFTALIAIDFQLSLEEFKSVGKVKETMLSPFEKLAQSLKETFSQAVQAMMSVTELR 240
QY 241 TNLNALKELITENTVTIQLKADAEQVRFLEKRRKQEEACSTLSHSIATLQESTT 300
DB 241 TNLNALKELITENTVTIQLKADAEQVRFLEKRRKQEEACSTLSHSIATLQESTT 300
QY 301 LKDDSTTNLHVESRLIGVMVQDGAESTVVEASQDDSAQPDQNSDAGEHKDS 355
DB 301 LKDDSTTNLHVESRLIGVMVQDGAESTVVEASQDDSAQPDQNSDAGEHKDS 355

RESULT 2

A72108

hypothetical protein CP0581 [imported] - Chlamydomophila pneumoniae (strains CWL029 and

C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae

C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000

C:Accession: A72108; B81560

R;Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: A72108
A:Molecule type: DNA
A:Residues: 1-390 <ARN>
A:Cross-references: GB:AE001605; GB:AE001363; NID:g4376455; PIDN:AAID8339.1; PID:g437645
A:Experimental source: strain CWL029
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: B81560
A:Molecule type: DNA
A:Residues: 1-390 <REA>
A:Cross-references: GB:AE002217; GB:AE002161; NID:g7189493; PIDN:AAF38399.1; PID:g718949
A:Experimental source: strain AR39, HL cells
A:Genetics:
A:Gene: CPn0186; CP0581

Query Match 11.5%; Score 200; DB 2; Length 390;
Best Local Similarity 23.0%; Pred. No. 0.00015;
Matches 87; Conservative 64; Mismatches 142; Indels 86; Gaps 13;

Qy 8 TSPVISRASSPTFGDHGKDFDNKKIIPISIEAPTSSAAVGAATAIEPGRSPLLQICY 67
Db 2 SSPVNTSPAPN-----IPIAPTTP-----KPRSSFIEKVI- 38
Qy 68 LVKIIAATLVVGGIAALVCLYLGVSIS-----TPSL-ILMLAIVSVFVITAIRDGT 121
Db 39 ---IVAKYILFAI---AATSGALGTILGSGALTPGIGIALLVFFVSWLLGLILKDSI 92
Qy 122 PSQVVRHMKQIOQFGEEN-----TRLHTAVENLKAVN----- 154
Db 93 SGGERRLRREVSRTSENORLTVITTTLETKVLDKAAKDQLTLEAFRNGENLKT 152
Qy 155 -VELSEQINOLKQLHRLSDFGRLEANTGDTFALLADQLSLEPKS-VGTVETMLSP 212
Db 153 AEDLEQVSKLSEQLALERINQLIQANAGDAQEISSELKLLISGWDKSVVEQINTSI-- 210
Qy 213 FEKLAQSLKETFSQAVQAMSSVTELTNLNALKELI-----TENKTIVIEQLKADAQ 265
Db 211 -----QALKVLLGQEWQEAQTHVKAMQEQIQALQAEILGMHNSOTALOKSVENLLVODQ 265
Qy 266 LREEQVRFLKRRKQLEAEACSTLHSHIATLQESTTLKDSITNLHVESRLIGVMVODGA 325
Db 266 ALTRVVGELLENKLSQACSLARQETIEKLAQHETSLOQRIDAMLAQEOQL-----A 317
Qy 326 ESSTVEASQDSDSAQPODE 344
Db 318 EQVTALERMKQEAQKAASE 336

RESULT 4
H81707
Inclusion membrane localised protein Inca TC0396 [imported] - Chlamydia muridarum (st
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
A:Accession: H81707
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255
A:Accession: H81707
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-276 <GT>
A:Cross-references: GB:AE002306; GB:AE002160; NID:g7190428; PIDN:AAF39253.1; PID:g719
A:Experimental source: strain Nigg (MoPn)
A:Genetics:
A:Gene: TC0396

Query Match 10.7%; Score 185; DB 2; Length 276;
Best Local Similarity 23.4%; Pred. No. 0.00075;
Matches 71; Conservative 64; Mismatches 121; Indels 48; Gaps 11;

Qy 16 SSPTFGDHGKDFDNKKIIPISIEAP--TSSAAVGAATAIEPGRSPLLQICYLVKIIA 73
Db 2 TSPTLVE-----MPLSCYPPTHTSTACTKRSSLYKPSLIETVQVAAVSL-- 49
Qy 74 AIALPVVGGIAALVCLYLGVSITPSLIILMLAIVSVFVITAIRDGTSPQVVRHMKQOI 133
Db 50 -AIIISIIIFAILGHAIFLAPQVALVIAVFIISLGNALYLCKTAPLRLYKELQOE 108
Qy 134 QOFGENTRLHTAVENLKAVNVSEQLNQLKHLTRISDFGRLEANTGDTFALLADQ 193
Db 109 ASLKEVNFEL-----LKSQKEF---LGLSKDFATSKDLSL-----VSLDFHNLQDDQ 154
Qy 194 LSLEEFKSVGTVETMLSPFEKLAQSLKETFSQAVQAMSSVTELTNLNALKEL----- 249

Db 155 SSHQGF-----EDLEDYKNSAEDLRQIFSOETVQSLSKSTILSKKEIKETVPTTEV 207
QY 250 --ITENKTVIEQLKADQ--LREQVRFLEKQKOEACSTLSHSTAT--LOESTTLLKDS 305
Db 208 RLRAENKEDLLKIVQDLQIRDK----LRAEINNLSQASKTILSEQIASQIENEXLIYANI 263
QY 306 TTNL 309
Db 264 TRAL 267

RESULT 5
C71553
Probable inclusion membrane protein A - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis
Reference number: A71570; MUID:99000809
Accession: C71553
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-273 <ARN>
A:Cross-references: GB:AE001286; GB:AE001273; NID:g3328516; PIDN:AAC67710.1; PID:g332851
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: inca

Query Match 9.5%; Score 164.5; DB 2; Length 273;
Best Local Similarity 22.8%; Pred. No. 0.011;
Matches 67; Conservative 59; Mismatches 109; Indels 59; Gaps 10;

QY 32 IIPISIEAPTSAAVGAKTAEPEGRSPLQRLICVLKIIAIALVFGIALLVCLYL 91
Db 8 VTPSPAPSYSANRV-----PQSLMDKIKKIAAASLTIGTIGFLALLGLHVG 58
QY 92 SVITSPSLIMLAIMVSVFVITAIRDTPSQVVR----HMKQIQOQFGEENTRLHTAV 147
Db 59 FLIAPQITIVLLAFITSLA-----GNALVLQKTANLHLQDQOR-----EV 100
QY 148 ENLKAVAVNELSEQINOLKQLTRLSDFGDLREANTGDTALADFQLSLEEFKSVGTKE 207
Db 101 GSLKEINFMLS--VLOKEFLH-----LSKEFATTSKDLASVQDFYSCLQGRDNYKGE 153
QY 208 TMLSPFEKLAQSLKETFSQBAVQAMSSVTELRITNLNALKELITENK-----TVI 257
Db 154 SLLDEYKNSTPEMKRFLSBOEIIADLKGVSASLREERFLTPAEVRRRAHNOQSLTVVI 213
QY 258 EQLKA-DAQLREB--QVRLEK-----RKQLEEEACSTLSHSTATQESTT 300
Db 214 EELKTRDSURDEIGOLSQSLKTLTSQIALQRKESDLCQIRETLLSPRKSAS 267

RESULT 6
G72593
Hypothetical protein APE1216 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix
Reference number: A72450; MUID:99310339
Accession: G72593
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-533 <KAW>
A:Cross-references: DBJ:AP0000061; NID:g5104821; PIDN:BAA80205.1; PID:g1043991; PID:g510

A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1216

Query Match 8.6%; Score 149; DB 2; Length 533;
Best Local Similarity 24.0%; Pred. No. 0.2; Mismatches 59; Conservative 46; Indels 109; Gaps 7;

QY 124 QVVRHMKQIQOQFGEENTRLHTAVENLKAVNVELSEQINOLKQLTRLSDFGDLREANTG 183
Db 251 QAVEDGLALDLSLEERVGDLFEFAVEDL---TLQLSLDSRSRGALEDVRADIEGRLEAVEG 307
QY 184 DF-----TALIADFQLSLEEFKSVGTGKVTMLSPPEKLAQSLKETFSQ--EAVQAMMS 234
Db 308 SLEDLSGAVDMSQQLAALAEDESLSRVEDL-----EARGSVEDRLSQAEEIDSLTT 363
QY 235 SVTELRITNLNALKELITENKTVITPOLKADQOLREEQVRFLEKQKOELEACSTLSHSTAT 294
Db 364 SLDSLRTELEDLSRLAEQAQASLEDLNRLDOVASTLQQLQRLATAEESLQALTEDLAS 423
QY 295 LOESTTLLKDDSTTNLHVESRLIGVMVQDGAESTVE-----EASODDSAQPDENQSD 348
Db 424 LOAEVETLQOSIVE---IDRRLL-----CQLRSTVDVAVRLEVESLGEKLVQAEKNQRQ 473
QY 349 AGEHKD 354
Db 474 DASIED 479

RESULT 7

T14039
protein kinase (EC 2.7.1.37), myotonic dystrophy-associated - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jun-2000
R:Leung, T.; Chen, X.Q.; Tan, I.; Manser, E.; Lim, L. Mol. Cell. Biol. 18, 130-140, 1998
A:Title: Myotonic dystrophy kinase-related Cdc42-binding kinase acts as a Cdc42 effector
Reference number: Z17862; MUID:98078670
Accession: T14039
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1732 <LEU>
A:Cross-references: EMBL:AF021935; NID:g2736150; PID:g2736151; PIDN:AAC02941.1
C:Genetics:
A:Gene: MRCK

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hem
C:Keywords: ATP; phosphotransferase
F:75-343/Domain: protein kinase homology <KIN>
F:1013-1062/Domain: protein kinase C zinc-binding repeat homology <KZN>

Query Match 8.4%; Score 145.5; DB 2; Length 1732;
Best Local Similarity 21.1%; Pred. No. 1.3;
Matches 54; Conservative 59; Mismatches 114; Indels 29; Gaps 7;

QY 126 VRHMKQIQOQFGEENTRLHTAVENLKAVNVELSEQINOLKQLTRLSDFGDLREANTGDF 185
Db 506 VNHLEQQLEEANSVRRELDADFQIKAFQIKTLOOEELNELKELVQASERLNQSKEL 565
QY 186 TALIADFQLSLEEFKSVGTGKVTMLSPPEKLAQSLKETFSQBAVQAMSSVTELRITNL-- 243
Db 566 KDAHQRKLAQMEFMEINERITELHTQKQLARVRO--KEEVEDLVQMAESLRQLR 623
QY 244 --NALKELITENKTVITPOLKADQOLREEQVRFLEKQKOELE---EACSTLSHSTATL--- 295
Db 624 AERAKKEVHTEALIAEASKDRKLREQSRHYSKOLENLEGLKQKQIISYSPGICSIHQ 683
QY 296 QESTTLLKD-----STTNLHVESRLIGVMVQD--GAESTVEE--ASODDS 338
Db 684 QEITKATDLEKKSIFVEEISREGIHAEINLKKELHDSGQQLANKEINVLKDKL 743

A:Reference number: S44243
A:Accession: S44243
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-254; C: 256-257, 'LQ', 260-276, 'A', 278-283, 'A', 285-519, 'D', 521-574, 'EQ', 5
A:Cross-references: EMBL:X78998; NID:9475933; PIDN:CAA55632.1; PID:9475934
C:Genetics:
A:Gene: GDB:EEA1
A:Cross-references: GDB:1369996
A:Superfamily: human early endosome antigen 1
C:Keywords: calmodulin binding; endocytosis; metal binding; peripheral membrane prote

Query Match 8.3%; Score 143; DB 1; Length 1410;
Best Local Similarity 24.0%; Pred. No. 1.4;
Matches 64; Conservative 52; Mismatches 99; Indels 52; Gaps 11;

QY 129 MKQIQOQFEE-----NTRLHTAVENLKAVNVSELSQINQLKQLATRLSDFGDRLEA 180
Db 812 LKQDFETLSQETKIQHEELNNRIQTTVTTELQVKKEKALMTLSVVKDKSKVSDSLKN 871
QY 181 NTGDF-----TALIADFQLSLEEFK-SVGTKVTMLSPFFKLAQSL-KETFSQEAQVA 231
Db 872 SKSEFEKENQKGAAILDLEKTKELKHQLOVQMENTLKEQELKKSLKEKEKASHQK 931
QY 232 MMSVTE-----LRTNLNALKELITENKTVIEQLKADAOALREEQVR 273
Db 932 ELNMOEQLIOAONTLKONEKEQQLQGNINELKQSEQKKQIEALQELKIAVLQKTE 991
QY 274 LE-KRKOELFEACSTLS---HSIATLQ-----ESTLLKDSSTNNLHVESRLIGVYQDGA 325
Db 992 LENKLOQLTQAQALAAEKEKISVLQNNYKESQETFKQLQSDFYGRESELLATR-----Q 1047
QY 326 ESSTVEE---ASQDD--SAQPDENQS 347
Db 1048 DLKSVVEKLSLAQEDLLISNRNQIGNQN 1074

RESULT 10
S26710
spindle pole body protein NUF1 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein D9476.3; protein YDR355W
C:Species: Saccharomyces cerevisiae
C>Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 23-Mar-2001
C:Accession: S26710; S34288; A49455; S61152
R:Marzavan, C.; Copeland, C.S.; Snyder, M.
J. Cell Biol. 116, 1319-1332, 1992
A:Title: The NUF1 gene encodes an essential coiled-coil related protein that is a pot
A:Reference number: S26710; MUID:92176232
A:Accession: S26710
A:Molecule type: DNA
A:Residues: 1-944 <MIR>
A:Cross-references: EMBL:Z11582; NID:94069; PIDN:CAA77668.1; PID:94070
A:Note: the authors translated the codon GAG for residue 206 as Asp and CTG for resid
R:Kilmartin, J.V.; Dyos, S.L.; Kershaw, D.; Finch, J.T.
submitted to the EMBL Data Library, June 1993
A:Description: A spacer element in the Saccharomyces cerevisiae spindle pole body who
A:Reference number: S34288
A:Accession: S34288
A:Molecule type: DNA
A:Residues: 1-944 <KIL>
A:Cross-references: EMBL:X73297; NID:9312173; PID:9312175
R:Kilmartin, J.V.; Dyos, S.L.; Kershaw, D.; Finch, J.T.
J. Cell Biol. 123, 1175-1184, 1993
A:Title: A spacer protein in the Saccharomyces cerevisiae spindle pole body whose tra
A:Reference number: A49455; MUID:94064779
A:Accession: A49455
A:Molecule type: DNA
A:Residues: 1-22 <KIL2>
A:Cross-references: EMBL:X73297
R:Du, Z.
submitted to the EMBL Data Library, June 1995
A:Description: The sequence of S. cerevisiae cosmid 9476.

QY 339 AQPDENQSDAGEBKD 354
Db 744 EKTRRESQSEREFEN 759

RESULT 8
T13030
microtubule binding protein D-Clip-190 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13030
R:Lantz, V.A.; Miller, K.G.
J. Cell Biol. 140, 897-910, 1998
A:Title: A class VI unconventional myosin is associated with a homologue of a microtubul
A:Reference number: Z17588; MUID:98139549
A:Accession: T13030
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1690 <LAN>
A:Cross-references: EMBL:AF041382; NID:92773362; PID:92773363; PIDN:AAB96783.1
A:Experimental source: strain Oregon R
C:Genetics:
A:Cross-references: FlyBase:FBgn0020503
C:Keywords: cytoskeleton

Query Match 8.4%; Score 145; DB 2; Length 1690;
Best Local Similarity 20.6%; Pred. No. 1.3;
Matches 64; Conservative 57; Mismatches 99; Indels 90; Gaps 10;

QY 126 VRHMKQIQOQGEENTRLHTAV-----ENLKAVNVEL-----SEQINQLKQL 167
Db 1057 IKNLQEEVTKAKTENLSTGTQTIKDLQERLEITNAELQHKEMASDAQKIADLKT 1116
QY 168 -----HTRLSDFGRLEANTGDTALIDFQSL-----EE 198
Db 1117 VEAIOVANANISATNAELSTVLEVLQAEKSETNHIFELFEMEADMNSERLIEKVTGKEE 1176
QY 199 FKSQTKVETMLSPFEKLAQSLKE-----TFQBAVQAMSSVTEL 239
Db 1177 LKETHQLDEROKKFEELKLAQSQEQKLOQESQTSKEKLTETIQSLQELQDSVVKQK 1236
QY 240 RTNLNALKELITENKTVIE-----QLKDAQ-LREQVRFLEKQOE--LEAC 285
Db 1237 EELVQNLEKVRSSIIIEAQNATKLNESNVQLENKTSCKETQDQLLESQKKEKOLQEEA 1296
QY 286 STLSHSIATLQESTTLTKDSTTNLH-----AVESRLIGVMVQDGAESVTEE-----ASQ 335
Db 1297 AKLSGELQVQVQANGDIKDSLKVVEELVKVLEEKLOAATSQLDAQAATNKELQELLVKSQ 1356
QY 336 DQSAQPDEN 345
Db 1357 ENEGNLOGES 1366

RESULT 9
A57013
early endosome antigen 1 - human
N:Alternate names: endosome-associated protein
C:Species: Homo sapiens (man)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A57013; S44243
R:Mu, F.T.; Callaghan, J.M.; Steele-Mortimer, O.; Stenmark, H.; Parton, R.G.; Campbell,
J. Biol. Chem. 270, 13503-13511, 1995
A:Title: EEA1, an early endosome-associated protein. EEA1 is a conserved alpha-helical p
A:Reference number: A57013; MUID:95286647
A:Accession: A57013
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1410 <RES>
A:Cross-references: GB:L40157; NID:g1016367; PIDN:AAA79121.1; PID:g1016368
R:Seeligm, H.P.
submitted to the EMBL Data Library, April 1994

A;Reference number: S61148
A;Accession: S61152
A;Molecule type: DNA
A;Residues: 1-944 <DUZ>
A;Cross-references: EMBL:U7
C;Genetics:
A;Gene: SGD:NUF1; SPC110
A;Cross-references: SGD:S00
A;Map position: 4R
C;Keywords: coiled coil; m

Query Match 8.1%; Score 141; DB 2; Length 944;
Best Local Similarity 23.5%; Pred. NO. 1.1;
Matches 61; Conservative 56; Mismatches 101; Indels 42; Gaps 8;

Qy	129	MKQIQOQEGENTRLHTAVENLK-----AVNVELSEQINOLKOLHTRLSDFGDR	177
Db	316	MDLQKQKQESKRKLDELNELETKFSENGSQSSAKENELMKLKNIAELEBEEISTKNSQ	375
Qy	178	LEANTQFTALIAIDFQLSREEFK-----SVGTKVETMLSPFPBKLAQSLK-----EFTFSO-	226
Db	376	LIAEGKSLASLA--QLTQLESKLNQDSQSGREBELAKTNDKLOKDRIAREETVSKD	433
Qy	227	EAVQAMSSVTELRTNLNALKELITENKTV-----IEQLKADQALREEQVRFLE	275
Db	434	ERIIDQKKVQKQENDLFDVFKKTHSESKTTIDNELESKDKLIKLENDLKVAQEKYKME	493
Qy	276	KRKQLEAEACSTLHSHSTATQESTTLLKQSTTNLNLHVESRLGVMVQDGAESTVVEEASQ	335
Db	494	K---ELKEREFNFKYESKLEDEKTTNEKISLNLAENSOLKNKI-----EDNSTATHMK	546
Qy	336	DDSAQQDQDNQSDAGEHKOS	355
Db	547	ENYEKQLESRLRKDIEEYKES	566

RESULT 11

outer membrane protein - Thermotoga maritima (strain MSB8)
H72216 C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: H72216
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.
C.W.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316
Accession: H72216
Status: preliminary
Molecule type: DNA
A:Residues: 1-403 <ARN>
A:Cross-references: GB:AE001812; GB:AE000512; NID:g4982302; PID:g498230
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1729

Query Match	8.1%	Score 140.5;	DB 2;	Length 403;
Best Local Similarity	21.0%;	Pred. NO. 0.44;		
Matches 73; Conservative	75;	Mismatches 113;	Indels 87;	Gaps 15;

Qy 47 VGAKTAPEGRSPLQRIICVLKIIIAATAALFVVGIAALVCLYLSVITPSILILM 106
Db 40 LGVLTGM-PDG---TFQGNSSLTRYQAQAVFY-----RNLKQPSADYVSLIN 85
Qy 107 LVSEFV---IVTATIR-----DGTPSQV-----VRHWK-----QQIQQ 135
Db 86 KVSITLEDLVSTATMKVQLNSDNFGGYTSDETLETKNDVANLKAATVLDKLNLRVEVMSQVQS 145
Qy 136 FGEE---NTRLLTAVENLKAVNVELSEQI-----NQLKQHTLRSLDFGDRL---EAN 181

Db	146	QSD	E	L	S	D	L	S	D	A	K	V	N	E	A	L	S	K	I	A	L	E	S	K	S	G	D	F	V	N	K	D	Y	V	D	S	K	I	A	Q	T	V	S	K	L	S	D	L	E	G	R	L	S	A	V	E	T	K	205
Qy	182	T	G	D	F	T	A	L	I	A	D	F	O	L	S	I	L	E	E	K	S	V	G	T	K	V	E	T	M	L	S	P	F	E	K	L	A	O	S	L	A	E	T	S	Q	E	A	V	M	M	S	S	V	T	E	L	T	241	
Db	206	T	A	N	E	A	L	V	R	N	S	A	S	K	D	Y	-----	V	E	K	L	S	Y	T	D	T	L	D	Q	K	L	S	E	L	S	A	S	-----	V	E	K	248																	
Qy	242	N	L	N	A	K	E	L	I	T	E	N	K	T	V	I	E	O	L	K	A	D	A	Q	L	R	E	E	E	R	F	L	E	K	R	Q	E	E	A	C	S	T	L	S	H	I	A	T	L	O	E	S	T	T	L	301			
Db	249	N	N	T	A	L	S	E	I	G	N	L	K	V	L	S	K	L	S	D	L	E	T	T	Q	O	K	T	A	R	D	A	L	D	A	R	V	S	L	E	G	I	T	T	V	N	S	R	S	L	E	K	R	V	S	Q	308		
Qy	302	L	K	D	S	T	T	N	L	H	A	V	E	S	R	L	I	G	Y	W	O	D	G	E	S	T	V	E	A	S	Q	D	S	T	V	E	A	S	Q	D	S	Q	D	S	N	S	D	A	349										
Db	309	V	E	S	A	D	K	V	N	S	L	E	-	R	M	G	A	V	T	---	A	R	V	T	K	V	E	-----	E	V	K	N	L	O	S	N	A	345																					

RESULT 12

F82383
methyl-accepting chemotaxis protein VCA1056 [Imported] - Vibrio cholerae (strain N1659)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: F82383
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
Chardon, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: F82383
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-770 <HEI>
A:Cross-references: GB:AE004432; GB:AE003853; NID:g9658509; PIDN:AAF96950.1; GSPDB:EN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VCA1056
A:Map position: 2

Query Match

```

Query Match      8.1%; Score 140; DB 2; Length 770;
Best Local Similarity 21.6%; Pred. No. 1;
Matches 73; Conservative 60; Mismatches 139; Indels 66; Gaps 13;

QY 68 LVKIIAIALFVVGIAALVCLYLSGVISTP--SLILMLAIMLVSEFVIVTAIRDGTp--- 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 417 IMKVLLIIVICAV-IATLAAMVWGNGISKPRDSISOIQMSRDNDETIVRLSEQSGSEIR 475
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 123 -----SOVRHMKOOIQOFGEE-----NTRLHTAVENLKAIVNVELSQ-----IN 162
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 476 QLAQALNMLSHLQDTICQFAHTTDKLNHSHTQTITHNMTGTRNSVSQEHRTDSVVTAVN 535
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 163 QLKQLHLRLSDFGDRL-----EANTGDETAIADFOQLSLEEFKSVGTVKVTMLSPPEKLA 217
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 536 EMSASITSEVSEFQARAATFVQEAENKQHGGSVSGNLA-RDMTSIN---QOMASAVEAIA 591
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 218 QSLKETTSQEAQVQAMSSVTELRNLNALKELLTENKT-----VIGQLKADAQ 265
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 592 RLNHESQSIALVDVIOQIAE-QTNLALNALNAEARAGEQGRGFVAVADEVRNLAAKTQ 650
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 266 LREQVRF-----LEKRQE-----LEEACSTLSHSIATLQESTTLLKDTSTNLHVESRLI 317
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 651 TSTBEIRTKIDRLQKETQSVVNCIEEANNIVRGVATCHNTDMLKQIVDMLNELNMNI 710
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 318 -----GVMVQDGAESTVEEASQDDSAQPDEN 345
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 711 QIATATEQQRGVTEENINANTISISDSVASVTVQGVGN 748
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 13

G02157
 kinesin-like spindle protein HKSP - human
 C:Species: Homo sapiens (man)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
 C:Accession: G02157

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 09:21:19 ; Search time 53.52 seconds
(without alignments)
736.756 Million cell updates/sec

Title: US-09-673-763-8

Perfect score: 1733

Sequence: 1 MTVSTDNTPVISRASSPTF.....DDSAQPQDENSAGEHKDS 355

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

A_Geneseq_032802.*

- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1733	100.0	355	20 AAY32173	Chlamydia psittaci
2	200	11.5	397	20 AAY34783	Chlamydia pneumoniae
3	162.5	9.4	273	20 AAY32176	Chlamydia trachoma
4	149	8.6	612	22 AAB95546	Human protein sequ
5	145	8.4	1690	22 ABB61144	Drosophila melanog
6	145	8.4	1690	22 ABB61173	Drosophila melanog
7	143.5	8.3	1286	21 AAB43359	Human ORF3123
8	143.5	8.3	2053	22 AAB43359	Human protein kina
9	143	8.3	1411	17 AAW02258	Nucleolar/endosoma
10	140.5	8.1	434	21 AAB42352	Human ORF ORF2116
11	140	8.1	575	22 AAB47215	Human KSP-S553. S

12	140	8.1	1057	22 AAG67419	Amino acid sequenc
13	140	8.1	1057	22 AAB47212	Human KSP. Homo s
14	139.5	8.0	739	22 AAB96493	Putative sensory t
15	138.5	8.0	344	21 AAB42170	Human ORF ORF1934
16	136	7.8	1325	18 AAW19540	Male-enhanced anti
17	136	7.8	1325	20 AAW94391	Mouse male enhance
18	134	7.7	475	22 ABB68543	Drosophila melanog
19	133.5	7.7	1948	22 ABB68543	Novel human diagno
20	133	7.7	880	22 AAB96332	Putative P. abyssal
21	131	7.6	286	18 AAW20404	H. pylori secreted
22	131	7.6	286	18 AAW24652	H. pylori secreted
23	131	7.6	327	18 AAW20901	H. pylori secreted
24	131	7.6	2482	16 AAW2826	Human mitotin amin
25	131	7.6	2482	19 AAW23996	Kinetochore protel
26	131	7.6	3248	17 AAR99795	Human secreted pro
27	130.5	7.5	945	20 AAW73624	Arabidopsis thalia
28	130.5	7.5	1521	21 AAG39235	Arabidopsis thalia
29	130.5	7.5	1528	21 AAG39234	Arabidopsis thalia
30	130.5	7.5	1562	21 AAG39233	Arabidopsis thalia
31	130.5	7.5	1703	21 AAG36714	Arabidopsis thalia
32	130.5	7.5	1710	21 AAG36713	Arabidopsis thalia
33	130.5	7.5	1728	21 AAB42089	Arabidopsis thalia
34	130.5	7.5	1744	21 AAG36712	Arabidopsis thalia
35	130.5	7.5	1797	22 AAG13880	Novel human diagno
36	130.5	7.5	2779	22 ABB62371	Staphylococcus aur
37	130	7.5	1184	22 AAG33733	Staphylococcus aur
38	130	7.5	1188	22 AAG33733	Staphylococcus aur
39	129.5	7.5	808	22 AAG05140	Novel human diagno
40	129.5	7.5	1483	21 AAG48640	Arabidopsis thalia
41	129.5	7.5	1493	21 AAG48639	Arabidopsis thalia
42	129.5	7.5	1544	21 AAG48638	Arabidopsis thalia
43	129.5	7.5	2633	22 AAG06505	Novel human diagno
44	129	7.4	1553	22 AAG02200	Novel human diagno
45	129	7.4	1788	22 AAG06749	Novel human diagno

ALIGNMENTS

RESULT 1
AAY32173
ID AAY32173 standard; Protein; 355 AA.
XX AC
XX AAY32173;
XX AC
XX 01-FEB-2000 (first entry)
XX DT
XX DE Chlamydia psittaci infection-specific protein Inca.
XX KW Inca; Infection; vaccine; therapy; diagnosis.
XX OS Chlamydia psittaci.
XX PN WO9953948-A1.
XX PD 28-OCT-1999.
XX PF 20-APR-1999; 99WO-US08744.
XX PR 20-APR-1998; 98US-0082438.
XX PR 21-APR-1998; 98US-0082588.
XX PR 22-MAY-1998; 98US-0086450.
XX (UYOR-) UNIV OREGON STATE.
XX Rockey DD, Bannantine JP;
XX WPI; 1999-633904/54.
XX N-PSDB; AAZ34587.
XX Novel bacterial infection specific proteins for treating and diagnosing
XX chlamydial infections

PS Claim 5; Page 39-40; 56pp; English.

XX This sequence represents novel infection-specific protein Inca of

CC Chlamydia psittaci strain GPIC. Inca is found in the inclusion

CC membrane of infected cells. It is associated primarily with the

CC vegetative reticulate body form of Chlamydia rather than with the

CC refractile elementary body form. The invention includes: a vaccine

CC directed against the reticulate body form of Chlamydia comprising 1

CC or more infection-specific proteins (see AAY32170-78), including

CC Inca, IncaB and IncaC; methods of using and producing such a vaccine;

CC methods for detection of infection-specific antibodies or antigens

CC in a biological specimen; and a method of using therapeutic agents

CC specifically directed against infection-specific peptides, or the

CC genes that code for such peptides, to treat chlamydial infection.

XX Sequence 355 AA;

SQ

Query Match 100.0%; Score 1733; DB 20; Length 355;

Best Local Similarity 100.0%; Pred. No. 4.4e-133;

Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTVDSTNTSPVTSRASSPTFGDHGKDFDNNKIIPISIEAPTSSAAAVGAKTAIEPEGGRSP 60

DB 1 mtvdstntspvtsrassptfgdhgkdfndnnkiipisieaptssaaavgaktaiepegrsp 60

QY 61 LLQRICYLKIIITAAIALFVVGIAALVCLYLGSVISTPSSLILMLATMLVSVIVITAIRDG 120

DB 61 llqricylkiiitaalffvgiaalvclylgsvistpsslilmlatmlvsvivitairdg 120

QY 121 TFSQVVRHMKQIOOQGEENTRLHVAENLKAVNVSELSQINQLKQLHTRLSDFGDRLEA 180

DB 121 tpsqvvrhmkqiooqgeentrlhvaenlkavnvselsqinqlkqlhtrlsdfgdrlea 180

QY 181 NTGDFITALIADFOLSELPKSVGTVKVTMLSPFEKLAQSLKETFSQEAQVQAMMSSVTELR 240

DB 181 ntgdfitaliadfolselepksvgtkvvtmlspfeklaqslketfsqeaqvqammsvte 240

QY 241 TNLNALKELITENKTVIEQLKADAOLREEQVRFLKRRKQLEAEACSTLSHSIATLQESTT 300

DB 241 tnlalnkelitenkvtieqlkadaolreeqvrflkrrkqleaeacstlshsiatquestt 300

QY 301 LLKDSSTNLHVESRLIGVMVQDGAESSTVEASQDSDSAQPDENOSDAGEHKDS 355

DB 301 llkdstnlhavesrligvmvqdgaaesstveasqdsdsagpdendgsagehkds 355

ULT 2

ID AAY34783

AC AAY34783;

XX 13-SEP-1999 (first entry)

DE Chlamydia pneumoniae transmembrane protein sequence.

XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

KW vaccine; neutralising epitope.

XX Chlamydia pneumoniae.

OS WO9927105-A2.

PN 03-JUN-1999.

XX 20-NOV-1998; 98WO-1B01890.

PF 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

XX (GEST) GENSET.

XX Griffiths R;

PI WPI; 1999-357842/30.

DR Genome sequence of Chlamydia pneumoniae

XX Page 760-761; Disclosure; 1912pp; English.

PS AAY34584-V35879 represent the proteins encoded by all the open reading

CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.

CC Chlamydia pneumoniae causes respiratory disease such as pneumonia and

CC bronchitis and is thought to be a contributing factor in heart

CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema

CC nodosum or pharyngitis. The polypeptides encoded by the open reading

CC frames of the C. pneumoniae genome (see AAY34584-V35879) can be used in

CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae

CC nucleotide sequences can also be used as immunogenic compositions,

CC especially where the vector directs the expression of a neutralising

CC epitope of C. pneumoniae.

XX Sequence 397 AA;

SQ

Query Match 11.5%; Score 200; DB 20; Length 397;

Best Local Similarity 23.0%; Pred. No. 4.7e-08;

Matches 87; Conservative 64; Mismatches 142; Indels 86; Gaps 13;

QY 8 TSPVISRASSPTFGDHGKDFDNNKIIPISIEAPTSSAAAVGAKTAIEPEGGRSP 67

DB 9 sspvntpsapn-----ipapttp-----giptp-----kprssfielvi- 45

QY 68 LVKIIIAIAIAFVVGIAALVCLYLGSVIS-----TPSL-ILMLAIAVSVIVITAIRDGT 121

DB 46 ---lvakiylfai---aatsgalgclilgsgaltcpigiallviffvsmvllgilkdsi 99

QY 122 PSQVVRHMKQIOOQGEENTRLHVAENLKAVN-----TLHPTAVENLKAVN----- 154

DB 100 sggeerlreevrsftsenqrlvtittletevkdkaakdqltleafrnengnltt 159

QY 155 -VEUSEQINQLKQLHTRLSDFGDRLEANTGDTALIAADFOLSELEFES-VGTRKVTMLSP 212

DB 160 aedleeqvskiseqlealerinqanagdaqeiselskklisgwskvveqints-- 217

QY 213 FEKLAQSLKETFSQEAQVQAMMSSVTELRNLNALKELI-----TENKTVIEQLKADAQ 265

DB 218 -----qalkvllgqevvqeaqthvkaqeqialqaeilgmhngstalgksvenllvqdq 272

QY 266 LREEQVRFLKRRKQLEAEACSTLSHSIATLQESTTLLKDSSTNLHVESRLIGVMVQDGA 325

DB 273 altrvvgellesenklqsacsalrqeleklaghsclgqridamlagaeln-----a 324

QY 326 ESSTVEASQDSDSAQPD 344

DB 325 eqvtalekmkqeaqaese 343

RESULT 3

AAY32176

ID AAY32176 standard; Protein; 273 AA.

XX AAY32176;

AC AAY32176;

XX 01-FEB-2000 (first entry)

DE Chlamydia trachomatis infection-specific protein Inca.

XX Inca; infection; vaccine; therapy; diagnosis.

XX Chlamydia trachomatis.

OS WO9953948-A1.

PN

XX

00 126 VRHMKQIQQFGEENTRLHTAV-----ENLKAVNVEL-----SEQINQKQL 167

QY 199 F K S V G T K V E T M L S P F E K L A Q S L K E - - - - - I F S Q E A V Q A R A D S V A L L E

Db 1057 iknlqeetvaktlenlelstgtgttkldqlerleithnaelqhkemasedaqkiadkl 1116
QY 168 -----HTRLSPDGRLEANTGDTTALADQLSL-----EE 198
Db 1117 vealqvanaisatnaelstvlvldqeksethifelfeemadmnserliekvgtikee 1176
QY 199 FKSQVTKVETMLSPFEKLAQSLKE-----TFQEAQVQAMWSSVTEL 239
Db 1177 lkethqliderqkfkfeeleklkqagseqklqgesqtskektelqgqlqelqdsvkqk 1236
QY 240 RTLNALKELITENKTVIE-----OLKADAO-LREQVRPLEKRRQE--LEEAC 285
Db 1237 eelqvleekvressiieaqtlnknesvnlentksciketqgdqlesqkqkqleq 1296
QY 286 STLSHIAIQLQESTLLKDDTNLH----AVESRLTGVWVQDGAESSTVEE-----ASQ 335
Db 1297 akisgelqvgqangdikdkslvkveelvleeklqaatsqldaqatnkelqellvksq 1356
QY 336 DQSAQPDQDN 345
Db 1357 enegnlqges 1366
RESULT 7
AAB43359
ID AAB43359 standard; Protein; 1286 AA.
XX AAB43359;
AC AAB43359;
DT 08-FEB-2001 (first entry)
XX Human ORFX ORF3123 polypeptide sequence SEQ ID NO:6246.
DE Human: open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antiporiatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive.
OS Homo sapiens.
XX WO200058473-A2.
XX 05-OCT-2000.
XX 31-MAR-2000; 2000WO-US08621.
XX 31-MAR-1999; 99US-0127607.
XX 02-APR-1999; 99US-0127636.
XX 05-APR-1999; 99US-0127728.
XX 30-MAR-2000; 2000US-0540763.
XX (CURA-) CURAGEN CORP.
XX Shinkets RA, Leach M;
XX WPI; 2000-602362/57.
XX N-PSDB; AAC77568.
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
PS Claim 11: Page 5433-5436; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40337 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antiporiatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 1286 AA;
Query Match 8.3%; Score 143.5; DB 21; Length 1286;
Best Local Similarity 24.3%; Pred. No. 0.0099;
Matches 58; Conservative 47; Mismatches 85; Indels 49; Gaps 8;
QY 127 RHMKQO---IQQFGENTRLHTAVENLKAIVNELSQI-----NQLKQLHRLS 172
Db 74 rmnkqgemseirrqkfyetqagklesqnrklesqkshqshdsknrilletrlr 133
QY 173 DFGDRLEANTGDTTALADQLSLSEEFKSVGTVKVTMLSPFE-KLAQS---LKETF--SQ 226
Db 134 evsleheqklelqkltelqslqeresqqltaqaaalesqrlrqakteleetaeae 193
QY 227 EAVQAMWSSVTELRYNLNALKKE---LITENKTVIFQLKAD----- 263
Db 194 eelqatahrdeiqkrfdalrnsctvitdleeqlnqltednaelnqnfyiskqideasg 253
QY 264 -----AQLREOVRELEKRRQKELEACSTLSHSIATLOESTTLKDDTNLHVESRLI 317
Db 254 andeivqlrse-vdhrlreiteremqltsqkqmealkttctmleeqvmdleandell 311
RESULT 8
AAU03501
ID AAU03501 standard; Protein; 2053 AA.
XX AAU03501;
XX 12-SEP-2001 (first entry)
XX Human protein kinase #1.
DE Human; protein kinase; PTK; STK; cancer; cardiovascular disease;
KW metabolic disorder; immune related disease; neurological disorder;
KW neurodegenerative disorder; inflammatory disorder; infectious disease;
KW reproductive disorder.
XX Homo sapiens.
OS
XX WO200138503-A2.
XX 31-MAY-2001.
XX 22-NOV-2000; 2000WO-US32085.
XX 24-NOV-1999; 99US-0167482.
XX (SUGE-) SUGEN INC.

XX PF 27-APR-1995; 95DE-1015514.
XX PR 27-APR-1995; 95DE-1015514.
XX PA (PRIV-) PRIVATES INST IMMUNOLOGIE & MOLEKULARGEN.
XX PI Renz M, Seelig HP;
XX DR WPI; 1996-403153/41.
XX DR N-PSDB; AAT58751.
XX PT DNA encoding nucleolar-endosomal auto-antigen - useful for exact
PT diagnosis of rheumatic disease, in gene therapy and for removal of
PT specific auto-antibodies
XX PS Claim 1; Fig 2; 15pp; German.
XX CC Transformed cells can be cultured to produce the antigen p162, for use
CC in exact (differential) diagnosis of rheumatic disease, i.e. they
CC can detect, in immunoassays, Western blots, etc., rheumatism-
CC specific auto-antibodies. The antigen can be used therapeutically,
CC in the removal of auto-antibodies from the circulation, or when
CC coupled to a cytotoxin, the elimination of auto-antibody-
CC producing lymphocytes.
XX SQ Sequence 1411 AA;
Query Match 8.3%; Score 143; DB 17; Length 1411;
Best Local Similarity 24.0%; Pred. No. 0.012;
Matches 64; Conservative 52; Mismatches 99; Indels 52; Gaps 11;
QY 129 MKQIQOQFEE-----NTRLHTAVENLKAVNVSEIQINLKOLHTRLSDFGDRLEA 180
Db 812 lkqdfetlsqetkiqheelnrriqtvtelqkvkmealmtelstvkdklskvsdskn 871
QY 181 NTGDF-----TALIADFOLSLEEFK-SVGTKEVTMLSPFEKLAOSL-KETFSQAVOA 231
Db 872 sksefekenqkaalidlektckelkhqlqvqmentlkeqkelkkslekeashqkl 931
QY 232 MMSSVTE-----LFTNLAKELITENKTVEIOLKADAOLREQVRF 273
Db 932 elnsmqeqligaqutlkqnekeeqqlqgninelkqseqkqkqlealqgelkviqkte 991
QY 274 LE-RRKQLEACSTLS---HSIATLQ-----ESTTLKDDSTTNLHAVESKLIGVMVDGA 325
Db 992 lenklqqltqaqaelaekesivlqnnnyeksgqetfkqlqsdqfygresellatr----q 1047
QY 326 ESSTVEE---ASQDD---SAQPQDENQS 347
Db 1048 dlksveekslslagedllsnrnqgnqn 1074
RESULT 10
AAB42352
ID AAB42352 standard; Protein; 434 AA.
XX AC AAB42352;
XX DT 08-FEB-2001 (first entry)
XX DE Human ORFX ORF2116 polypeptide sequence SEQ ID NO:4232.
XX KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
PD 12-SEP-1996.
XX PI Plowman GD, Whyte D, Manning G, Sudarsanam S, Martinez R;
XX PI Flanagan P, Clary D;
XX DR WPI; 2001-343950/36.
XX DR N-PSDB; AAS06701.
XX PT Nucleic acids encoding human kinase polypeptides, useful for preventing
PT diagnosing and/or treating e.g. cancer, immune, cardiovascular and
PT neuronal-associated diseases, and microbial infections -
XX PS Claim 7; Figure 2; 433pp; English.
XX CC AAU03501-AAU03557 represent novel human protein kinases #1-57. The
CC novel protein kinases have been identified as members of the tyrosine
CC or serine/threonine kinase (PTK and STK) families. The polynucleotides
CC encoding protein kinases and the polypeptides may be used in the
CC prevention, diagnosis and treatment of diseases associated with
CC inappropriate kinase expression. For example, they may be used to treat
CC cancers (especially cancers of haematopoietic origin), cardiovascular
CC disease (e.g. atherosclerosis), metabolic disorders (e.g. diabetes),
CC immune related diseases (e.g. rheumatoid arthritis), neurological
CC disorders (e.g. schizophrenia), neurodegenerative disorders (e.g.
CC Parkinson's disease), inflammatory disorders (e.g. asthma), infectious
CC disease (e.g. HIV) and reproductive disorders (e.g. infertility).
CC Additionally, polynucleotides encoding protein kinases may be
CC used for gene therapy and as DNA probes in diagnostic assays.
CC The protein kinase polypeptides may be used as antigens in the production
CC of antibodies against the protein kinases and in assays to identify
CC modulators of protein kinase expression and activity.
XX SQ Sequence 2053 AA;
Query Match 8.3%; Score 143.5; DB 22; Length 2053;
Best Local Similarity 24.3%; Pred. No. 0.019;
Matches 58; Conservative 47; Mismatches 85; Indels 49; Gaps 8;
QY 127 RHMKQO---IQFGENTRLHTAVENLKAVNVSEIQI-----NOLKQLHTRLS 172
Db 843 rnmkaqeemselrqqkfyelqagkqlaqnkrkleekiekishqhsdknrlleltr 902
QY 173 DFGDRLEANTGDTALIADFOLSLEEFKSVGTKEVTMLSPFP-KLAQS---LKETF--SQ 226
Db 903 evsleheeqklelkrqltelqslqeresqltalqaaraalesqlrqakteleetaae 962
QY 227 EAVQAMSSVTELTNLALKE---LITENKTVEIOLKAD----- 263
Db 963 eeiqaltahrdelqrkfdalrnsctvitdeeqinqdtednaelnqnfylskqldaeag 1022
QY 264 ----AQLREEQVRLEKRRQLEEEACSTLSHSIATLQSTTLKDDSTTNLHAVESRLI 317
Db 1023 andeiqlrse-vdhrirreiteremqltsqkqmealktctmleeqvmdlealndell 1080
RESULT 9
AAW02258
ID AAW02258 standard; Protein; 1411 AA.
XX AC AAW02258;
XX DT 09-MAR-1997 (first entry)
XX DE Nucleolar/endosomal auto-antigen p162.
XX KW Auto-antibody; p162; rheumatic disease; antigen; diagnosis;
KW gene therapy.
XX OS Homo sapiens.
XX PN DE19515514-C1.
XX PD 12-SEP-1996.

cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive.

Homo sapiens.

WO200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000WO-US08621.

31-MAR-1999; 99US-0127607

02-APR-1999; 99US-0127636.
05-APR-1999; 99US-0127728

30-MAR-2000; 2000US-0540763.

(CURA-) CURAGEN CORP.

Shimkets RA, Leach M.

WPI; 2000-602362/57.
N-PSDB: AAC76551

Verbal analogies

useful for treating

Claim 11; Page 3424-3425; 5507pp; English.

AAC74446 to AAC77606 encode the proteins a

sequences have activities such as: cytostatic; hepatoprotective; vulnery; antispasmodic; antiparkinsonian; nootropic; immunoprotective; immunostimulant; cardiant; antiarthritic; immunopressant; antidiabetic; hypotensive; dermatolytic; coagulant; vasotropic; antiinflammatory; antibacterial; dermatological; immunosuppressive; antihydroid; and antianemic. The sequences of antifungal, antirheumatic, antithyroid, and antitubercular. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy. The proteins can be used to express ORFX proteins in gene therapy. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, ocular haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

sequence 434 AA;

Query Match	8.1%	Score 140.5;	DB 21;	Length 434;
Best Local Similarity	26.4%;	Pred. No. 0.0038;		
Matches	58;	Conservative 32;	Mismatches 77;	Indels 53;
				Gaps 8;

[illegible]

	QY	184	DFTALIA-----DFQLSLEEFKSVCTKVETMLS-----pf-----	213
			: : :: : :	:

01 u v s m a v e e y e e m q v n l e k d r k k a e s f a q e m f l e p n g g k k t p p f g r q s s i l d q q l 140

QY 214 -----E K L A Q S I K E T F S O E A V Q A M S S V T E L R T N L N A L K E L I T E N K T V --- I E O L K A D A Q 265

QY 266 LREEQVRFLEKRRQLEEEACSTLSHSTATLQESTLLKDS 305
| | | | : | : | | | | | : : |
Db 188 lleeedkkelclkygnseekarnlkhsvdelqkrvngns 227

RESULT 11

AAB47215
ID AAB47215 standard; Protein; 575 AA.
....

AC AAB47215;

18-JUL-2001 (first entry)

Human KSP-S553.

Human; kinesin; KSP; drug screening; mitotic spindle; mitosis; ATP hydrolysis; apoptosis; necrosis; cancer

Synthetic.

AX
PN
WO200131335-A2.

03-MAY-2001

26-OCT-2000; 2000WO-IIS29570

27-OCT-1999: 99US-0428156

(CYTO-) CYTOKINETICS INC.

Wood KW, Flner JT, Beraud C, Mak J, Sakowicz R;

R WPI; 2001-300550/
R N-PSDB; AAC85783.

Screening for drug candidates and treating cancer by using kinesin

Disclosure; Fig 8; 63pp; English.

This sequence represents the human kinesin, KSP-S553, which may be used in the method of the invention to screen for drug candidates and bioactive agents. The method comprises combining the candidate and a cell expressing recombinant KSP, and determining the effect of the candidate on the bioactivity of recombinant KSP. Changes in the bioactivity of KSP may be determined by assays for determining changes in the mitotic spindle, particularly inhibition of mitosis, and ATP hydrolysis. It may also be determined by performing assays to determine the effect of candidate agents on apoptosis and necrosis. The method of the invention is useful for screening for drug candidates (especially bioactive agents and proteins) which effect the bioactivity of KSP, binding to KSP and/or expression of KSP, where the cells are cancer cells.

Sequence 575 AA:

Query Match	8.1%	Score 140;	DB 22;	Length 575;
Best Local Similarity	21.7%	Pred. No. 0.0062;		

Matches 56; Conservative 56; Mismatches 100; Indels 46; Gaps 113

```

|||||: || ||: | :: : | | : : : | : | : | : | :
298 vitalvertphvpyresktrllqdslg--grtrtsi---tatispaslnleestley 352
167 LHTPISDEGQDRIYVNG DMTF-----

```

353 ah-raknlnhpvngqltkkalikeyteerlrdlaaarekngvyiseenfrvmsqk 411

222 EFTSQEAVQAMSSVTELRNLNALKELTENKTIVIEQLKADAQREQVRFLEKKRQEL 281

412 ltvqeegivelliekigaveeelrvtelmdnknlelgcksallgnktqettkghlet 471

Matches 56; Conservative 50; MaxMatchScore 70

Qy	113	VITAIRDCTP-----SQVVRHMKIQOIFGEENTKLTAFTAVNLKAVNVFISEQINQLKQ	166
		: :: :: : :: : :: :	
Dh	298	Vitalvertphvpyresklrllqdslg-grtrsi---iatipasinleestlstley	352

Query Match			
Best Local Similarity	21.7%	Pred. No.	0.015;
			10;
		Indels	46; Gaps
			100;

QY 113 VITAIKDTGP-----SOVVRHMKQIQOFGENTRLHTAVENLKAVNVLSLSEQINQLK 166
 Db 298 litalvertphvpyreskltrildqsg--grtrtsi---iatispaslnleeltstley 352
 QY 167 LHTRLSDGDLREANTG-DFTALIDFQLSLEEFK----SVGTRKVTMLSPFEKLAQSLK 221
 Db 353 ah-raknllnkpevnqkltkalkikeyteelerikrdlaaarekngvyiseefrnmsgk 411
 QY 222 ETSQEAQVAMMSSVTELRNLNALKELITENKTVIOLKADQLREQVRFLEKRKQEL 281
 Db 412 ltvqeegivelliekigaveeelnrvtfmdnkneidqcksdignktqelctgkhlget 471
 QY 282 EACSTLSHATLOES--TTLKDKSTNHLHVESRLIGVMVDGAESSTVEEASOD--- 336
 Db 472 k-----lqivkeeyitsalesteeklhdaaskil-----ntveettkdvsg 512
 QY 337 -----DSAQPDQENQSDA 349
 Db 513 lnskldrrkkavdqhnae 530

RESULT 14
 ID AAB96493 standard; Protein; 739 AA.
 AC AAB96493;
 DT 29-OCT-2001 (first entry)
 DE Putative sensory transduction histidine kinase and response regulator #3.
 KW Hyperthermophilic archaeon; hyperthermophilic protein.
 OS Pyrococcus abyssi.
 PN FR2792651-A1.
 PD 27-OCT-2000.
 PF 21-APR-1999; 99FR-0005034.
 PR 21-APR-1999; 99FR-0005034.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 PA (IFRE-) IFREMER INST FR RECH EXPL MER.
 PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
 PI Querellou J, Weissenbach J, Saurin W, Hellig R;
 PI WPI; 2001-126236/14.
 PS New nucleotide sequences isolated from Pyrococcus abyssi encode
 proteins useful in industry -
 Claim 7; Pages 1203-1205; 1657pp; French.
 CC The present invention relates to the genomic sequence of Pyrococcus
 CC abyssi (see AAF86431 and AAF41223-7) and P. abyssi proteins. P. abyssi is
 CC a hyperthermophilic archaeon, which is isolated from deep-sea
 CC hydrothermal vents. The present invention is one such P. abyssi protein.
 CC The proteins of the present invention have various potential industrial
 CC uses, since the proteins are stable at very high temperatures, some up to
 CC 110 degrees centigrade.
 CC Note: This patent is in the same patent family as WO2000065062, which
 CC contains additional sequences as shown in AAB99132-AAB99143,
 CC AAB75903-AAH75920 and AAG66436.
 CC Sequence 739 AA.

Query Match 8.0%; Score 139.5; DB 22; Length 739;
 Best Local Similarity 21.1%; Pred. No. 0.0096;
 Matches 63; Conservative 66; Mismatches 124; Indels 45; Gaps 13;

QY 71 IIAIAL-FVWGIAALVCLY-LGSVISTPSLILMLAILVSVIVITAIKDTGPSQVVRH 128
 Db 308 viqgfavgfviilvavilviykiasntlap-----lekkyaaalaegrkqvsey 358
 QY 129 MKQOIQOFGENTRLHTAVENLKAVNVLSLSEQINQLKHTRLSDGDLREANT---GD 184
 Db 359 lk-qiryl--erdegaligafavskdlvgtlnaisklerlae-gdlsngltvevrg 414
 QY 185 FTALIDFQLSLEEFK-SVGTRKVTMLSPFEKLAQSLKETFQ--EAVQAMMSSVTELR 241
 Db 415 lrdilidirsvtetfreslgsive-mandlekranlaqvskdvtealngvnealqgvs 473
 QY 242 NLNALKEKITENKTVIEQLKADQLREQVRFLEKRKQELSEACS----- 286
 Db 474 eaqrqetene---itdgmrlvaqtseesvrameefsgavteevsvianegsqgdealkr 530
 QY 287 --TLSHSIATLQESTTLTKDSTNHLHVESRLIGVMVDG--AESSTVEEASQDSDAQ 340
 Db 531 lediqhmmrieetvskvaemsrnieitnvtisaeqtnllalnaaeaaeageagr 588

RESULT 15
 ID AAB42170 standard; Protein; 344 AA.
 AC AAB42170;
 DT 08-FEB-2001 (first entry)
 DE Human ORF1934 polypeptide sequence SEQ ID NO:3868.
 KW Human; open reading frame; ORF; detection; cytostatic; hepatotropic;
 KW vulnary; antipariatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 OS Homo sapiens.
 PN WO2000058473-A2.
 PD 05-OCT-2000.
 PF 31-MAR-2000; 2000WO-US08621.
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 PA (CURA-) CURAGEN CORP.
 PI Shimkets RA, Leach M;
 PI WPI; 2000-602362/57.
 DR N-PSDB; AAC76379.
 XX Novel nucleic acids and peptides derived from open reading frame X,
 XX useful for treating e.g. cancers, proliferative disorders,
 XX neurodegenerative disorders and cardiovascular disease -
 PS Claim 11; Page 3015-3016; 5507pp; English.

Tue Aug 13 14:23:42 2002

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective; anticonvulsant; antiarthritic; immunosuppressant; osteopathic; anticonvulsant; thrombolytic; coagulant; vasotropic; immunostimulant; cardiant; dermatological; immunosuppressive; antidiabetic; hypotensive; antibacterial; antiviral; antifungal; antirheumatic; antiinflammatory; antianaemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX proteins in gene therapy. Nucleic acids can be used to express ORFX proteins in gene therapy. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

CC Sequence 344 AA:
XX
SQ

Query Match		8.0%;	Score 138.5;	DB 21;	Length 344;
Best Local Similarity		22.8%;	Pred. No. 0.004;		
Matches 59;		Conservative 52;	Mismatches 115;	Indels 33;	Gaps 8;
QY	114	ITAIRDGTSPQVVRH---MKQIQQFGENTRLHTAVENLKAVNVEL-----SEQINQ	163		
Db	87	ltalkgalkkeevshdqemdklkeqydaelqalresvee-aknveviasrntseqdqa	145		
QY	164	LKQLHTR-LSDFGDRLAEANTGDTALIADPQLSLEEFKSVGKVTETMLSPPEKLAQSLKE	222		
Db	146	gtemrvkllqeeneklqgrseeleerrvaqlqrqiedlkqdeakaketikkyegeirqllee	205		
QY	223	TF-----SOEAVQAMSSVTELTNLNALKELITENKTVIEQLKADAOLEEQVRFLEK	276		
Db	206	alvharkeekeavsarraleeneleaagqnlsgttqeqqlseklkeeseqeqq---lrr	261		
QY	277	RKQLEECSTLSHSTATLQ-ESTTLKDKDSTTNLHAVESRL-----IGVMVQDGA	325		
Db	262	lknemenerwhlgtkieklkemadiveasrtstlelqnlqdeykeknrrrelaemqrlk	321		
QY	326	ESSTVEEASQDSDSAQPODE	344		
		l : : : : : lll			
	322	ektlaeeksriltankmqde	340		

Search completed: August 13, 2002, 09:23:46
Job time: 147 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 09:24:24 ; Search time 17.41 Seconds
(without alignments)
789.514 Million cell updates/sec

Title: US-09-673-763-8

Perfect score: 1733

Sequence: 1 MTVSTDNTPVSRSSPTF.....DSSAQPDENQSDAGEHKDS 355

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	143.5	8.3	1286	1	CTRO_HUMAN
2	141	8.1	944	1	NUF1_YEAST
3	140	8.1	1057	1	EGS_HUMAN
4	139	8.0	866	1	MYSP_SCHMA
5	139	8.0	1805	1	HMW2_MYCGE
6	138.5	8.0	1597	1	CTRO_MOUSE
7	136.5	7.9	1940	1	MYH3_HUMAN
8	136.5	7.9	1940	1	MYH3_RAT
9	136	7.8	1325	1	GLI6_MOUSE
10	135.5	7.8	1337	1	MYH8_HUMAN
11	134	7.7	1790	1	USO1_YEAST
12	133.5	7.7	1939	1	MYH4_HUMAN
13	133	7.7	880	1	RA50_PYRAB
14	132	7.6	433	1	HTR2_HALVA
15	132	7.6	539	1	MY53_HYDAT
16	132	7.6	1957	1	YD86_SCHPO
17	131.5	7.6	516	1	PS4_ENTFC
18	131.5	7.6	993	1	SCP1_MOUSE
19	131	7.6	1969	1	MYSA_CAEEL
20	131	7.6	3210	1	CENF_HUMAN
21	131	7.6	3911	1	AKA9_HUMAN
22	130	7.5	886	1	RA50_SULAC
23	129.5	7.5	2663	1	CENE_HUMAN
24	129	7.5	1972	1	MYHB_HUMAN
25	128.5	7.4	473	1	YVCE_BAGSU
26	128.5	7.4	1938	1	MYH4_RABIT
27	128.5	7.4	1939	1	MYH1_HUMAN
28	128	7.4	2022	1	ANT1_ONCVO
29	127.5	7.4	2349	1	TPR_HUMAN
30	127	7.3	1938	1	MYHD_HUMAN
31	127	7.3	1972	1	MYHB_MOUSE
32	127	7.3	1976	1	MYHA_RAT
33	126.5	7.3	1084	1	MYSS_RABIT

34	126.5	7.3	1130	1	YL17_CAEEL
35	126.5	7.3	1972	1	MYHB_RABIT
36	126.5	7.3	1978	1	MYHB_CHICK
37	126	7.3	866	1	MYSP_SCHJA
38	126	7.3	882	1	RA50_PYRFU
39	126	7.3	1244	1	MIX1_CAEEL
40	126	7.3	1433	1	REST_CHICK
41	126	7.3	1976	1	MYHA_HUMAN
42	126	7.3	2245	1	MYSU_DICDI
43	125.5	7.2	1184	1	BIMC_Emeni
44	125.5	7.2	1966	1	MYSB_CAEEL
45	125	7.2	697	1	MFPI_LYCES

ALIGNMENTS

RESULT	1
CTRO_HUMAN	
ID	CTRO_HUMAN
AC	014578; Q9UP27; STANDARD; PRT; 1286 AA.
DT	30-MAY-2000 (Rel. 39, Created)
DT	30-MAY-2000 (Rel. 39, Last sequence update)
DT	01-MAR-2002 (Rel. 41, Last annotation update)
DE	Citron protein (Rho-interacting, serine/threonine kinase 21) (Fragment).
DE	CIT OR STK21 OR KIAA0949.
GN	Homo sapiens (Human).
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC	NCBI_TaxID=9606;
OX	[1]
RN	SEQUENCE FROM N.A.
RP	Connell M., Goela D., Harper M.;
RA	Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RL	[2]
RN	SEQUENCE OF 347-1286 FROM N.A.
RC	TISSUE=Brain;
RX	MEDLINE=99246063; PubMed=10231032;
RA	Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT	"Prediction of the coding sequences of unidentified human genes. XIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro."
RL	DNA RES. 6:63-70(1999).
CC	-!- FUNCTION: PUTATIVE RHO/RAC EFFECTOR THAT BINDS TO THE GTP-BOUND FORMS OF RHO AND RAC1. IT PROBABLY BINDS P21 WITH A TIGHTER SPECIFICITY IN VIVO (BY SIMILARITY).
CC	-!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC	-!- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG BINDING DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR	EMBL; AC002563; AB71327.1; -
DR	EMBL; AB023166; BAA76793.1; -
DR	MIM; 605629; -
DR	InterPro; IPR001180; CNH.
DR	InterPro; IPR002219; DAG_PE-bind.
DR	InterPro; IPR001849; PH
DR	InterPro; IPR000861; REM_repeat.
DR	Pfam; PF00780; CNH; 1.
DR	Pfam; PF00130; DAG_PE-bind; 1.
DR	Pfam; PF00169; PH; 1.
DR	SMART; SM00109; C1; 1.
DR	SMART; SM00036; CNH; 1.
DR	SMART; SM00233; PH; 1.

RA Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ESSENTIAL COMPONENT OF THE NUCLEOSKELETON. POTENTIAL
CC ROLE IN CROSSLINKING FILAMENTS OR ANCHORING OTHER MOLECULES. IT
CC IS ESSENTIAL FOR GROWTH.
CC -!- SUBCELLULAR LOCATION: NUCLEAR; TIGHTLY ASSOCIATED WITH THE
CC NUCLEUS. IT IS PRESENT IN A GRANULAR PATTERN THAT EXCLUDES THE
CC NUCLEOLUS.
CC -!- PTM: MAY BE REGULATED BY PHOSPHORYLATION EVENTS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z11582; CAA77668.1; -
CC EMBL; X73297; CAA51733.1; -
CC EMBL; U28372; AAB64791.1; -
CC PIR; S26710; S26710.
CC PIR; S34288; S34288.
CC SGD; S0002764; NUF1.
CC Coiled coil; Nuclear protein; Phosphorylation.
CC KW DOMAIN 164 791
CC FT DOMAIN 54 59
CC FT DOMAIN 726 731
CC FT DOMAIN 742 747
CC FT DOMAIN 731 944
CC FT DOMAIN 731 944
CC SEQUENCE 944 AA; 111781 MW; 04FAA074BB8A0BC8 CRC64;
CC
Query Match 8.1%; Score 141; DB 1; Length 944;
Best Local Similarity 23.5%; Pred. No. 0.45;
Matches 61; Conservative 56; Mismatches 101; Indels 42; Gaps 8;
QY 129 MKOOIOQFGENTRLHTAVENLK-----AVNVELSEIOLKQLHTRLSDFGR 177
DB 316 MDLQKQKONESKRLDELNELETKFSENGSQSSAKENELKMLKNAIELEIEISTKNSQ 375
QY 178 LEANTGDTALADFQLSLEEFK-----SVGTQKVTMLSPFPEKLAQSLK----ETFSQ- 226
DB 376 LIAKEGKLASLMA--QUTQLESKLNQDSQSGREELKTKNDKQDIARIAREETVSKD 433
QY 227 EAVQAMSSVTELTNLNALKELITENKV-----IEOLKADQLREOVRFLE 275
DB 434 ERIDIDKKVKQLENDLFVIKTKHSEKTTIDNELESKDKLILKLENDLKVAQEKYSKME 493
QY 276 KRQOELEACSTLSHSTATQESTTLTKDSTNLHVESRLIGVMVODGAESSTVEEASQ 335
DB 494 K---ELKREFNFYKISEKLEDEKTTLNKISLNLAENSQLNKNI-----EDNSTATHMK 546
QY 336 DDSAQPDENOSDAGEHKDS 355
DB 547 ENYEKQLESRLKDIIEYKES 566
RESULT 3
EG5_HUMAN
ID EG5_HUMAN STANDARD; PRT; 1057 AA.
AC P52732; O15716;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Kinesin-related motor protein Eg5 (Kinesin-like spindle protein HKSP)
DE (Thyroid receptor interacting protein 5) (TRIP5)
DE 1).
GN KNSL1 OR EG5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

DR PROSITE; PS00479; DAG_PE_BIND_DOM.1; 1.
DR PROSITE; PS0081; DAG_PE_BIND_DOM.2; 1.
DR PROSITE; PS0003; PH_DOMAIN.1.
KW Coiled coil; Phorbol-ester binding; SH3-binding.
FT NON_TER 1
FT DOMAIN <1 558
FT DOMAIN 365 561
FT DOMAIN 509 512
FT DOMAIN 622 670
FT DOMAIN 702 822
FT DOMAIN 851 1148
FT SITE 1212 1217
FT SEQUENCE 1286 AA; 146506 MW; 498101F79EA75E85 CRC64;
CC
Query Match 8.3%; Score 143.5; DB 1; Length 1286;
Best Local Similarity 24.3%; Pred. No. 0.47;
Matches 58; Conservative 47; Mismatches 85; Indels 49; Gaps 8;
127 RHMKQK---IQQFGENTRLHTAVENLKAVNVELSEI-----NOLKQLHTRLS 172
DB 74 RNKMQEEMISLQKQFYLETQAGLEAQNRLKEQLEKISHQDSDKNRLLELETRLR 133
QY 173 DFGRLQANTGDTALADFQLSLEEFKSVGTQKVTMLSPFE-KLAQS---LKETF--SQ 226
DB 134 EVSLEHEEQKLELRQLTELQSLQRESQITLQAAARAALESQLRQAQTELETTAAE 193
QY 227 EAVQAMSSVTELTNLNALKELITENKVIEQKAD-----LITENKVIEQKAD----- 263
DB 194 BEIQALTARHDEIQKFDALNRKCTVITDLEQLNQLTQEDNAELNNQNFYLSKOLDEASG 253
QY 264 -----AQIREQVFLERKQLEENAGSTLSHSTATQESTTLTKDSTNLHVESRLI 317
DB 254 ANDEIVQLRSE-VDLHRLREITERMQLTQSQKOTWEALTKTCTMLEEQVMDLEALNDELL 311
RESULT 2
NUF1_YEAST
ID NUF1_YEAST STANDARD; PRT; 944 AA.
AC P32380;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE NUF1 protein (Spindle body spacer protein spc110).
DE NUF1 OR SPC110 OR YDR356W OR D9476.3.
OS Saccharomyces cerevisiae (Baker's yeast).
OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OS Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
[1]
SEQUENCE FROM N.A.
RN STRAIN=S288C;
RC MEDLINE=92176232; PubMed=1541631;
RX Mirzayan C., Copeland C.S., Snyder M.;
RA "The NUF1 gene encodes an essential coiled-coil related protein that
RT is a potential component of the yeast nucleoskeleton."
RL J. Cell Biol. 116:1319-1332(1992).
RN [2]
SEQUENCE FROM N.A.
RN STRAIN=S288C / AB972;
RC MEDLINE=94064779; PubMed=7503995;
RX Kilmartin J.V., Dyos S.L., Kershaw D., Finch J.T.;
RA "A spacer protein in the Saccharomyces cerevisiae spindle body
RT whose transcript is cell cycle-regulated."
RL J. Cell Biol. 123:1175-1184(1993).
RN [3]
SEQUENCE FROM N.A.
RN STRAIN=S288C / AB972;
RC Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favetto A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Rifken L., Riles L., Taich A., Trevaskis E., Vignati D.,

STABILIZES THE SHAPE OF THE WALL-LESS MYOPLASMA. THIS CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HW PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHESIN PROTEINS IN THE MYOPLASMA MEMBRANE AT THE ATTACHMENT ORGANELLE (BY SIMILARITY).

THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION - THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL ENTITIES REQUIRES A LICENSE AGREEMENT (SEE <http://www.isb-sib.ch/announce/> OR SEND AN EMAIL TO license@sib-sib.ch).

EMBL: U39701; AAC71437.1; .
EMBL: U02165; AAD12447.1; .
TI6R; MG218; .
DR InterPro; IPR003364; Seryl_tRNA_N; 1
DR Pfam; PF02403; Seryl_tRNA_N; 1
KW Cytadherence; Structural protein; Coiled coil; Complete proteome.
FT DOMAIN 28 838 COILED COIL (POTENTIAL).
FT DOMAIN 914 1591 COILED COIL (POTENTIAL).
FT DOMAIN 1632 1723 COILED COIL (POTENTIAL).
FT DOMAIN 1777 1804 COILED COIL (POTENTIAL).
FT SEQUENCE 1805 AA; 216252 MW; 11D093AF173284FD CRC64;
SQ

Query Match 8.0%; Score 139; DB 1; Length 1805;
Best Local Similarity 23.1%; Pred. No. 1.3; 103; Indels 90; Gaps 13;
Matches 71; Conservative 44; Mismatches

QY 116 AIDGTPSVVRHMKOIOQFGEENTRLHTAVE---NLK---AVNVSESEQINQLKQHLT 169
DB 304 SFODGTTKQNAQHVDEKLVALNKEKDRNTQKEAFENLRQSLIDINKLQENEL----- 358
QY 170 RLSDFGDRLEANTGDFITALIDFQLSLE-----EFKS-VGTVKVTMLSP---F 213
DB 359 ----FAKHLHQNEFEQKQSDLSLLKLETEYKALQHKINEFKNESATKSEELLNQERELF 414
QY 214 EK-----LAQ-SLKETFSQEAQVAMMSVTTELRTNLN-----ALKELTENTVTEQLK 261
DB 415 EKREIDTLTQASLEYEHQRESSQLLKQNEVKQHFQNLQYAKKELDKERNLLDQOKK 474
QY 262 ADA-----QLREOVRF---LEKRK 278
DB 475 VDSEAFQALKEKVAQERKELELYLVKKQKODQKENELFFKQKQHQADFENELEAKQ 534
QY 279 QLEEEACSTLSHSTATQESTTLKLDSTNLHVESRLIGVMVDGAESESTVEASQDS 338
DB 535 QELFEAKHALERSFIKLEDEK---KDLNTRQAQIANE-FSQKTKDKSKADFELMLQNEY 590
QY 339 AQPDQENQ 346
DB 591 ENLQKEQK 598

RESULT 6
CTRO_MOUSE
ID CTRO_MOUSE STANDARD; PRT; 1597 AA.
AC P49025;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Citron protein (Rho-interacting, serine/threonine kinase 21).
GN CIT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96128238; PubMed=8543060;

FT CONFLICT 423 428 DOVKEL -> GSSORI (IN REF. 2).
FT CONFLICT 430 431 SS -> KL (IN REF. 2).
FT CONFLICT 637 639 V -> L (IN REF. 2).
FT CONFLICT 639 639 T -> I (IN REF. 2).
FT CONFLICT 691 691 G -> E (IN REF. 2).
FT CONFLICT 720 720 S -> F (IN REF. 2).
FT SEQUENCE 866 AA; 100387 MW; 42FA5E6E78176AE0 CRC64;
SQ

Query Match 8.0%; Score 139; DB 1; Length 866;
Best Local Similarity 21.4%; Pred. No. 0.53; 93; Indels 86; Gaps 14;
Matches 67; Conservative 67; Mismatches

QY 126 VRHMKOIOQFGEENTRLHTAVEN-----LKAVNVSESEQINQLK-----QLHTRLSD- 173
DB 383 VNTLTSNLSOLESENLEKSLVNDLTDKNLLERENRQNDQVKELKSSLRDANRLTDL 442
QY 174 --FQDRLEANTGDFITALID-----FQLSLEEFKSVGTVKVTMLSPPEKLAQSL 220
DB 443 EALRSQLEARDNLASALHDAEEALHDMQYQASQAALNHLKSEMEQRLRERDEBESL 502
QY 221 KETSQEAQVAMMSVTTEL-----RTNLNALKE-----LITENK 254
DB 503 KKS--TTRTIEELTVITHEMEVKYSLSRLKKRYESNIADLEIQLODTANKANANLAKENK 561
QY 255 TVIEQLK-----ADAQLREEQVRFLE---KRKQ---ELEACSTL-----SHSIA 293
DB 562 NLSQVRKDLTEFLDEERLRREARNLQITEHKRLQLANEEIRSTLENLRKHAET 621
QY 294 TLOESTTLKSTNLHAV---ESRL---IGVMVDGAESESTVEASQD-----DSAQ 340
DB 622 ELEEAQSRVSELTQVNTLTNDKRLKLECDIGVQADMDDAINAQASQEDRAIRLNNEVLR 681
QY 341 PQDENQSDAGEHK 353
DB 682 LADELRLQEQGNK 694

RESULT 5
HMW2_MYCGE
ID HMW2_MYCGE STANDARD; PRT; 1805 AA.
AC P47460;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytadherence high molecular weight protein 2 (Cytadherence accessory protein 2).
HMW2 OR MG218.
Mycoplasma genitalium.
Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Ufferback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.F., Dougherty B.A., Bort K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 557-659 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bort K.F., Hutchison C.A. III;
RA "A survey of the Mycoplasma genitalium genome by using random
sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
RX -!- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH

```

RA Madanle P., Furuyashiki T., Reid T., Ishizaki T., Watanabe G.,
RT Morii N., Narumiya S.;
RL "A novel partner for the GTP-bound forms of rho and rac.";
RC FEBS Lett. 377:243-248(1995).
CC -1- FUNCTION: PUTATIVE RHO/RAC1 EFFECTOR THAT BINDS TO THE GTP-BOUND
CC FORMS OF RHO AND RAC1. IT PROBABLY BINDS P21 WITH A TIGHTER
CC SPECIFICITY IN VIVO.
CC -1- SUBUNIT: HOMODIMER (PROBABLE).
CC -1- TISSUE SPECIFICITY: A MAJOR SIGNAL WAS OBSERVED IN TESTIS AND
CC BRAIN, BUT IT WAS ALSO DETECTED IN THYMUS, SPLEEN, KIDNEY, HEART
CC AND LUNG.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U39904; AAC52341.1; .
CC MGD; MGI:105313; Cit.
CC DR InterPro; IPR001180; CNH.
CC DR InterPro; IPR002219; DAG_PE-bind.
CC DR InterPro; IPR001849; PH.
CC DR InterPro; IPR000861; REM_repeat.
CC DR Pfam; PF00780; CNH; 1.
CC DR Pfam; PF00130; DAG_PE-bind; 1.
CC DR Pfam; PF00169; PH; 1.
CC DR SMART; SM00109; C1; 1.
CC DR SMART; SM00036; CNH; 1.
CC DR SMART; SM00233; PH; 1.
CC DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 1.
CC DR PROSITE; PS00081; DAG_PE_BIND_DOM_2; 1.
CC DR PROSITE; PS00003; PH_DOMAIN; 1.
CC DR Colic coil; Phorbol-ester binding; SH3-binding.
CC FT DOMAIN 1 845 COILED COIL (POTENTIAL).
CC FT DOMAIN 674 870 RHO/RAC BINDING.
CC FT DOMAIN 818 821 POLY-IYS.
CC FT DOMAIN 931 979 PHORBOL-ESTER AND DAG BINDING.
CC FT DOMAIN 1011 1131 PH.
CC FT DOMAIN 1160 1457 CNH.
CC FT SITE 1521 1526 SH3-BINDING (POTENTIAL).
CC SQ SEQUENCE 1597 AA; 183448 MW; 7B7286C2305676DA CRC64;

Query Match 8.0%; Score 138.5; DB 1; Length 1597;
Best Local Similarity 23.8%; Pred. No. 1.2;
Matches 57; Conservative 48; Mismatches 85; Indels 49; Gaps 8;

QY 127 RHMKQ---IQFGENTRLHTAVENLKAVNVLSQI-----NQLQLHRLS 172
DQ 127 RHMKQEMISELQKFLYETQAGKLEQNRKLEQLEKISHQHSKRSLELETRLR 442
QY 173 DFGDRLEANTGDTALIAQLSLEEFKSVGTVKTMSPFE-KLAQS---LKETF--SQ 226
DQ 443 EVSLEHEEKLEKRLQLEQLSLQRESQLTALQAARALESQLRQAKTELEETAEAE 502
QY 227 EAVQAMSSVTELRNLNKAKE---LITENKTVIQLKAD----- 263
DQ 503 EETQALTAHRDEIQRFDAIRNSCVITDLEQLNLTEDNALNNQNFYLSKQLEASG 562
QY 264 -----AQLREQVRFLEKKRQBEACSTLSHSIATLQESTTLLKSDTTNTHAVESRLI 317
DQ 563 ANDEIVQLRSE-VDLHRLREITEREMQLTSQKQTEALKTCTMTLEEQVLDLEALNDELL 620

RESULT 7
MYH3_HUMAN
ID MYH3_HUMAN STANDARD; PRT; 1940 AA.

```

```

AC P11055; Q15492;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, fast skeletal muscle, embryonic (Muscle embryonic
DE myosin heavy chain) (SMHCE).
GN MYH3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89263803; PubMed=2726495;
RA Eller M.S., Stedman H.H., Sylvester J.E., Fertels S.H.,
RA Rubinstein N.A., Kelly A.M., Sarkar S.;
RT "Nucleotide sequence of full length human embryonic myosin heavy
RT chain cDNA.";
RL Nucleic Acids Res. 17:3591-3592(1989).
RN [2]
RP SEQUENCE OF 774-1940 FROM N.A.
RX MEDLINE=90033298; PubMed=280546;
RA Eller M.S., Stedman H.H., Sylvester J.E., Fertels S.H., Wu Q.-L.,
RA Raychowdhury M.K., Rubinstein N.A., Kelly A.M., Sarkar S.;
RT "Human embryonic myosin heavy chain cDNA. Interspecies sequence
RT conservation of the myosin rod, chromosomal locus and isoform
RT specific transcription of the gene.";
RL FEBS Lett. 256:21-28(1989).
RN [3]
RP SEQUENCE OF 856-1940 FROM N.A.
RX TISSUE=Skeletal muscle;
RC MEDLINE=90235862; PubMed=1691980;
RA Bober E., Buchberger-Seidl A., Braun T., Singh S., Goedde H.W.,
RA Arnold H.H.;
RT "Identification of three developmentally controlled isoforms of human
RT myosin heavy chains.";
RL Eur. J. Biochem. 189:55-65(1990).
RN [4]
RP SEQUENCE OF 856-1940 FROM N.A.
RX MEDLINE=89366648; PubMed=2771643;
RA Karsch-Mizrachi I., Travis M., Blau H., Leinwand L.A.;
RT "Expression and DNA sequence analysis of a human embryonic skeletal
RT muscle myosin heavy chain gene.";
RL Nucleic Acids Res. 17:6167-6179(1989).
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1- DEVELOPMENTAL STAGE: ABUNDANTLY PRESENT IN FETAL SKELETAL MUSCLE
CC AND NOT PRESENT OR BARELY DETECTABLE IN HEART AND ADULT SKELETAL
CC MUSCLE.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X13988; CAA32167.1; .

```

DR EMBL; X13100; CAA31492.1; -
DR EMBL; X51593; CAA35942.1; -
DR EMBL; X15696; CAA33731.1; -
DR PIR; S04090; S04090.
DR HSP; P08799; IMMD.
DR MIM; 160720; -
DR InterPro; IPR000048; IQ.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRODOM; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00096; IQ; 1.
DR Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
Calmodulin-binding; ATP-binding; Methylation; Alkylation;
Multigene family.
KW DOMAIN 1 781 MYOSIN HEAD-LIKE.
FT DOMAIN 782 811
FT DOMAIN 840 1933 COILED COIL (POTENTIAL).
FT NP_BIND 179 186 ATP (POTENTIAL).
FT DOMAIN 656 678 ACTIN-BINDING.
FT DOMAIN 758 772
FT MOD_RES 130 130 METHYLATION (SH-1).
FT MOD_RES 696 696 ALKYLATION (SH-2).
FT MOD_RES 706 706 A -> G (IN REF. 3).
FT CONFLICT 1331 1331 KK -> QE (IN REF. 1 AND 2).
FT CONFLICT 1391 1392 SR -> RA (IN REF. 3).
FT CONFLICT 1608 1609 RG -> QT (IN REF. 2).
FT CONFLICT 1663 1664
SQ SEQUENCE 1940 AA; 224035 MW; 43CA58C6A4BA1253 CRC64;

Query Match 7.9%; Score 136.5; DB 1; Length 1940;
Best Local Similarity 21.4%; Pred. No. 1.9; Indels 33; Gaps 7;
Matches 52; Conservative 56; Mismatches 102; Indels 33; Gaps 7;

QY 121 TPGSVVRHMQQ-----IQQFGEENTRLHTA-VENLKAVN-----VELSQINQLKOLHTR 170
Db 1158 TSTQIENKKREAFELKLRDLRDLEATLQHEAVATLKKHADSVAELGEQIDNLQKRVK-- 1215
171 LSDFGDRLEANTGFTALADFLQSLSEPKSVGKVTMTLSPFEKLAOSLKTFSQ---- 226
1216 -----QKLEKSEFKLEIDDLSSME-----SVSKSKANLEKICRTLEDQLSEARGK 1263
QY 227 -EAVQMMSSVT---ELRTNLNALKELITENKTVIQLKADAKQLREEQVRFLEKRRQEL 281
Db 1264 NEEQRSLSLELTOKSRLOTEAGELSRLQLEKESIVLSQSRSKQAFTQOTEELKRLQLEE 1323
QY 282 BEACSTLSHSIATQESTTLTKSTTNLHVESRLIGVMQDGAESSTVEASODDSAQP 341
Db 1324 NKAKNALAHQLSSRHDCDLLREQYEEQEGKAEQLRALSKANSEVAQWRTKYETDAIQR 1383
QY 342 QDE 344
Db 1384 TEE 1386

RESULT 8
MYH3_RAT ID MYH3_RAT STANDARD; PRT; 1940 AA.
AC P12847;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE MYosin heavy chain, fast skeletal muscle, embryonic.
GN MYH3.
OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87060988; PubMed=3783701;
RA Nadal-Ginard B.;
RA Streher E.E., Streher-Page M.-A., Perriard J.C., Periasamy M.,
RT "Complete nucleotide and encoded amino acid sequence of a mammalian
RT myosin heavy chain gene. Evidence against intron-dependent evolution
RL J. Mol. Biol. 190:291-317(1986).
CC -|- FUNCTION: MUSCLE CONTRACTION.
CC -|- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -|- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -|- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -|- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -|- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -|- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -|- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; X04267; CAA27817.1; -
CC PIR; A24922; A24922.
CC HSP; P08799; IMMD.
CC InterPro; IPR000048; IQ.
CC InterPro; IPR004009; Myosin_N.
CC InterPro; IPR002928; Myosin_tail.
CC InterPro; IPR001609; myosin_head.
CC Pfam; PF00612; IQ; 2.
CC Pfam; PF00063; myosin_head; 1.
CC Pfam; PF02736; Myosin_N; 1.
CC Pfam; PF01576; Myosin_tail; 1.
CC PRINTS; PR00193; MYOSINHEAVY.
CC PRODOM; PD000355; myosin_head; 1.
CC SMART; SM00015; IQ; 1.
CC SMART; SM00242; MYSC; 1.
CC PROSITE; PS00096; IQ; 1.
CC Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
KW Multigene family.
FT DOMAIN 1 781 MYOSIN HEAD-LIKE.
FT DOMAIN 782 811
FT DOMAIN 840 1933 COILED COIL (POTENTIAL).
FT NP_BIND 179 186 ATP (POTENTIAL).
FT DOMAIN 656 678 ACTIN-BINDING.
FT DOMAIN 758 772
FT MOD_RES 130 130 METHYLATION (SH-1).
FT MOD_RES 696 696 ALKYLATION (SH-2).
FT MOD_RES 706 706 A -> G (IN REF. 3).
FT CONFLICT 1331 1331 KK -> QE (IN REF. 1 AND 2).
FT CONFLICT 1391 1392 SR -> RA (IN REF. 3).
FT CONFLICT 1608 1609 RG -> QT (IN REF. 2).
FT CONFLICT 1663 1664
SQ SEQUENCE 1940 AA; 223857 MW; B5D546A5965A696 CRC64;

Query Match 7.9%; Score 136.5; DB 1; Length 1940;
Best Local Similarity 21.4%; Pred. No. 1.9;
Matches 52; Conservative 57; Mismatches 101; Indels 33; Gaps 7;

QY 121 TPQVVRHMQO-----IQOFGENTRLHTA-VENLAVN-----VELSEQINOLKQLHTR 170
DQ 1158 TSTQELNKKREAEFLRLDRLEATLQHEATVATLRKKHADSAELAEOQIDNLRKYK-- 1215
QY 171 LSDFGDLEANTGFTALIDFOLSEEFKSVCTKVTMLSPFEKLAQSLKETFSQ----- 226
DQ 1216 -----QKLEKESEFKLEIDDLSSV-----SVSKSKANLEKICRTLEDQLSEARGK 1263
QY 227 -EAVQAMSSVT-----ELRTNLNALKELITENKTVIOLKADAKLREEQVRFLEKQKQEL 281
DQ 1264 NEETQRLSELTKQSLQTEAGELSKOLEEKESIVSOLSRKSAFTQOIEELKQLEEE 1323
QY 282 BEACSTLSHAIQESTTLLKDDSTNLHAVESRLICVMVQDGAESSTVEEASQDSDAOP 341
DQ 1324 NKAKNALAHALQSSRHDCDLLREQYEEQBGKAEQLQALSKANSEVAQWRTKYETDAIOR 1383
QY 342 QDE 344
DQ 1384 TEE 1386

SULT 9

60_MOUSE
ID G160_MOUSE STANDARD; PRT; 1325 AA.
AC P55937;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Golgin-160 (Male-enhanced antigen-2) (MEA-2).
GN GOLGA3 OR MEA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1; TISSUE=Testis;
RX MEDLINE=97217683; PubMed=9063644;
RA Kondo M., Sutou S.;
RT "Cloning and molecular characterization of cDNA encoding a mouse
male-enhanced antigen-2 (Mea-2): a putative family of the Golgi
autoantigen.";
RL DNA Seq. 7:71-82(1997).
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN SPERMATOGENESIS AND/OR
TESTIS DEVELOPMENT. PROBABLY IDENTICAL WITH THE SEROLOGICALLY
DETECTABLE MALE ANTIGEN (SDM).
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTIS. TRANSCRIPTS CAN BE
FOUND IN SPERMATIDS DURING SPERMATOGENESIS. NO EXPRESSION IN
LEYDIG CELLS, SPERMATOGENIA, OR SPERMATOCYTES.
CC -!- SIMILARITY: HIGH, TO HUMAN GOLGIN-160.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-19 OR MET-30 IS THE
INITIATOR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; D78270; BAA19612.1; -
DR HSSP; P18852; ISCG.
DR MGD; MGI:96958; Golga3.
KW Spermatogenesis; Developmental protein.
FT DOMAIN 201 204 POLY-ALA.
SQ SEQUENCE 1325 AA; 149880 MW; 3230636962C687B0 CRC64;

Query Match 7.8%; Score 136; DB 1; Length 1325;
Best Local Similarity 23.6%; Pred. No. 1.3;
Matches 67; Conservative 38; Mismatches 78; Gaps 9;

QY 126 VRHMQOIQO-----FGEENTRL-----HTAV-----ENLKAVNVLSQIN----- 162
DQ 405 LOHARQWYQQOQLAQEARVRLQGEWAHQVQOMTQAGLLEHLKLENVSLSHQLTQHR 464
QY 163 -----QKQLHTRLS-----FGDLEANTGFTALIDFOLSEEFKSVCTKVTET 208
DQ 465 SIKERITAVQLQSTQSTADMLDQEAFAVQIREAK-----TWVEDLQRRLEEFEGEREQLOK 520
QY 209 MSLSPFEKLAOSLKE-----TFQBAVQAMSSVTETLRTN 242
DQ 521 VADAASLEQQLQEQVKLTLFORDQOALAQOEHLQVIRKQLSTOALQAKGSLDLDLHTR 580
QY 243 LNALKELITENKTVIOLKADAKLREEQVRFLEKQKQELACSTLSHAIQESTTLL 302
DQ 581 YDEL-----QARLEELQREADSREDAIHFLQNEKIVLEVALQSAKSKDELQRGARRL 633
QY 303 KDDSTNLHAVESRLICVMVQDGAESSTVEEASQDSDAOPDENQ 346
DQ 634 EEDTEE-----TSGLEQLRODLAVKSNQVHEHQQETATLRKQM 673

RESULT 10

MYH8_HUMAN
ID MYH8_HUMAN STANDARD; PRT; 1937 AA.
AC P13535; Q14910;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-JUN-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, skeletal muscle, perinatal (MyHC-perinatal).
GN MYH8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=90323631; PubMed=2373371;
RA Karsch-Wizrachi I., Feghali R., Shows T.B. Jr., Leinwand L.A.;
RT "Generation of a full-length human perinatal myosin heavy-chain-
encoding cDNA.";
RL Gene 89:289-294(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=95324556; PubMed=7601129;
RA Jullian E.H., Kelly A.M., Pompidou A.J., Hoffman R., Schiaffino S.,
Stedman H.H., Rubinstein N.A.;
RT "Characterization of a human perinatal myosin heavy-chain
transcript.";
RL Eur. J. Biochem. 230:1001-1006(1995).
RN [3]
RP SEQUENCE OF 502-1937 FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=90235862; PubMed=1691980;
RA Bober E., Buchberger-Seidl A., Braun T., Singh S., Goedde H.W.,
Arnold H.H.;
RT "Identification of three developmentally controlled isoforms of human
myosin heavy chains.";
RL Eur. J. Biochem. 189:55-65(1990).
RN [4]
RP SEQUENCE OF 860-1937 FROM N.A.
RX MEDLINE=89234168; PubMed=2715179;
RA Feghali R., Leinwand L.A.;
RT "Molecular genetic characterization of a developmentally regulated
human perinatal myosin heavy chain.";
RL J. Cell Biol. 108:1791-1797(1989).
RN [5]
RP SEQUENCE OF 1-46 FROM N.A.
RA Esser K., Tidhar A., Myszkowski M.;
RT "Isolation and characterization of the human perinatal MHC promoter.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MUSCLE CONTRACTION.

130 KOQIOQGE-----ENTRLHTAVENLKAYNVELSEGINOLKOLHTRLSDFGDRLEA----- 180
1076 KOOLDEKLEKKFEINLSKISTEDQAVETIQOKKI---KELQARIELGEEIEAERASR 1132
181 -----NTGFTALIAIDFQLSLEEFKSVGTQKVTML-----SPFKEKLAOSLKE-TFISOEAV 229
1133 AKAEKORSDLSELEIEISERLEE-AGGATSAQVELNKKREAEFOKLRDLLEATLQHEAM 1191
230 QA-----MMSSVTELTNTNALKELITENKTVIEOLKADAQLRREQVRFLEKRRQEL 281
1192 VAALRKKHADSMALGEEQIDNLRQVKQLEKES---ELKMETDSSLNAAEAIKAKGNL 1248
282 EEACSTLSHSIATL-----QESTTLTKDSTNTNLHAVESRLIGVMVDGASSTVEASODD 337
1249 EKMCRSLEDOVSELKTKEEEOQLINDLTAQRARLQTEAGEYSRQIDKDALVSQLSRSK 1308
338 SAQOP 342
1309 QASTQ 1313
RESULT 11
USOL_YEAST STANDARD; PRT; 1790 AA.
ID USOL_YEAST
AC P25386;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Intracellular protein transport protein USOL.
GN USOL OR INT1 OR YDL058W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=91185402; PubMed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RT Yamasaki M.;
RA "A cytoskeleton-related gene, usol, is required for intracellular
RT protein transport in Saccharomyces cerevisiae.";
RL J. Cell Biol. 113:245-260(1991).
RN [2]
RP SEQUENCE OF 782-1790 FROM N.A.
RA Hostettler M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA Kendrick K.E.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC COMPLEX.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC ER AND THE GOLGI COMPLEX.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC -!- SIMILARITY: BELONGS TO THE VOP/USOL/YBL047C FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; X54378; CAA38253.1; -

Query Match	7.8%	Score 135.5;	DB 1;	Length 19377;
Best Local Similarity	22.9%	Pred. NO. 2.2;		

DR EMBL; L03188; AAB00143.1; -;
DR EMBL; U53668; AAB66659.1; -;
DR PIR; A38455; A38455;
DR HSSP; P80220; 1DIP.
DR SGD; S0002216; USOL.
DR InterPro; IPR002017; Spectrin.
KW Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
FT DOMAIN 1 724 GLOBULAR HEAD.
FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
FT CONFLICT 847 847 G -> E (IN REF. 2).
FT CONFLICT 924 924 E -> K (IN REF. 2).
FT CONFLICT 1253 1253 V -> I (IN REF. 2).
FT CONFLICT 1319 1319 I -> V (IN REF. 2).
FT CONFLICT 1461 1461 N -> S (IN REF. 2).
FT CONFLICT 1581 1581 G -> S (IN REF. 2).
FT CONFLICT 1600 1600 I -> V (IN REF. 2).
FT CONFLICT 1661 1661 R -> S (IN REF. 2).
FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).
FT SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;
Query Match 7.78; Score 134; DB 1; Length 1790;
Best Local Similarity 21.38; Pred. No. 2.5;
Matches 57; Conservative 53; Mismatches 121; Indels 36; Gaps 8;
QY 117 IRDTPSQVVRHMKQIQOFGENTRLHTAVENLKAVNVE-----LSEQ 160
Db 1372 LNEGS-STIQEYSEKINTLEDELIRLQNEE-LKAKEIDNTRSELEKVSLSNDELLEEK 1429
QY 161 INQLQLHLRLSDGRLEANTGDTALADFDLSLEEFK-----SVGTVETMLSPF 213
Db 1430 QNTIKSLQDELISKYKIDTRNDEKLISIERDNKRDLESKEQLRAAQESKAKVEEGLKL 1489
QY 214 EKLAQSLKETF-----SQEAYQAMKSSVTELTNLNALKELITENKTVIQLKADAOLREE 269
Db 1490 EE-ESSKKAELKESKEMMKLESIESNETELKSMETIRKSDKLEQSKSA---EE 1544
QY 270 QVRFLEKKELEAEASTLSHSTATQESTTLTKDSTNLHAVESRLIGV--VODGAS 327
Db 1545 DINLQHEKSDLSIRNESEKDEELKSLRIEAKSGSELTVKQLNNAQEKIRINAE 1604
QY 328 STVEASQDASQPDENSDAGEHKD 354
Db 1605 NTVLKSLEIDIERELKDKQAEIKSNOE 1631

RESULT 12

MYH4_HUMAN
AC MYH4_HUMAN STANDARD; PRT; 1939 AA.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, skeletal muscle, fetal (Myosin heavy chain IIB)
DE (MyHC-IIB).
GN MYH4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=93318869; PubMed=10388558;
RA Wells A., Schiaffino S., Leinwand L.A.;
RT "Comparative sequence analysis of the complete human sarcomeric myosin
heavy chain family: implications for functional diversity";
RL J. Mol. Biol. 290:61-75(1999).
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2

CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HWM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AF111783; RAD29949.1; -;
CC HSSP; P08799; IMMD.
CC MIM; 160742; -;
CC InterPro; IPR000048; IQ.
CC InterPro; IPR004009; Myosin_N.
CC InterPro; IPR002928; Myosin_tail.
CC InterPro; IPR001609; myosin_head.
CC Pfam; PF00612; IQ; 2.
CC Pfam; PF00063; myosin_head; 1.
CC Pfam; PF02736; Myosin_N; 1.
CC Pfam; PF01576; Myosin_tail; 1.
CC PRINTS; PR00193; MYOSINHEAVY.
CC ProDom; PD000355; myosin_head; 1.
CC SMART; SM00015; IQ; 1.
CC SMART; SM00242; MYSC; 1.
CC PROSITE; PS0096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
KW Multigene family.
FT DOMAIN 1 784 MYOSIN HEAD-LIKE.
FT DOMAIN 785 814 IQ
FT DOMAIN 843 1939 COILED COIL (POTENTIAL).
FT NP_BIND 179 186 ATP (POTENTIAL).
SQ SEQUENCE 1939 AA; 223012 MW; 40B1AD1D77A47DE CRC64;

Query Match 7.78; Score 133.5; DB 1; Length 1939;
Best Local Similarity 20.4%; Pred. No. 2.9; Mismatches 55; Gaps 5;
Matches 52; Conservative 55; Mismatches 123; Indels 25; Gaps 5;
QY 121 TPSQVVRHMK-----QIQOFGENTRLH-----TAVENLKAVNVE-ELSEQINOLKOLHTR 170
Db 1161 TSAQIELNKKRAEFQKMRDLEESTLQHEATAAALRKKHADSVAEGLQKIDSLQVRKQK 1220
QY 171 LSDFGDRLEANTGDTALADFDLSLEEFKSVGTVETMLSPFLEKKELEAEASTLSH 230
Db 1221 LEKEKSELKNEINDLASNMETVSKANFEKMCRTLEDQLSEIK-----TKEEQORLIN 1275
QY 231 AMSSVTELTNLNALKELITENKTVIQLKADAQREQVRFLEKKELEAEASTLSH 290
Db 1276 ELQAQKARLHTESEGFQRDLDEKDMVSQLSRGKQAFQIQIELKQLEETKAKSTLAH 1335
QY 291 STATQESTTLTKDSTNLHAVESRLIGVWODGASSTVEASQDASQPDENQ-----346
Db 1336 ALQSAHDCDLLREQVEEPQEAELQKMGSKANSEVAQWRKYETDAIQTETEELEAKK 1395
QY 347 -----SDAGEHKDS 355
Db 1396 KLAQRLQDAEEHVEA 1410

ID	HTR2_HALVA	STANDARD;	PRT; 433 AA.
AC	P42258;		
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	15-DEC-1998 (Rel. 37, Last annotation update)		
DT	Sensory rhodopsin II transducer (HTR-II) (Methyl-accepting phototaxis protein II) (MPP-II) (Fragment).		
DE	HTRII.		
GN	Haloarcula vallismortis.		
OS	Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloarcula.		
OC	NCBI_TaxID=28442;		
OX	SEQUENCE FROM N.A.		
RP	STRAIN=ATCC 29715;		
RC	MEDLINE=95224074; PubMed=7708770;		
RX	Seidel R., Scharf B., Gautel M., Kleine K., Oesterheld D., Engelhard M.;		
RA	"The primary structure of sensory rhodopsin II: a member of an additional retinal protein subgroup is coexpressed with its transducer, the halobacterial transducer of rhodopsin II.";		
RT	proc. Natl. Acad. Sci. U.S.A. 92:3036-3040(1995).		
RL	-!- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR SENSORY RHODOPSIN II (SR-II) TO THE FLAGELLAR MOTOR. RESPONDS TO LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION.		
CC	-!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.		
CC	-!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).		
CC	EMBL; Z35308; CAA84549.1; .		
DR	InterPro; IPR004089; Chemotaxis_transducer.		
DR	InterPro; IPR003660; HAMP.		
DR	Pfam; PF00015; MCPsignal; 1.		
DR	SMART; SM00304; HAMP; 1.		
DR	SMART; SM00283; WA; 1.		
DR	Transducer; Photoreceptor; Transmembrane; Methylation.		
KW	NON_TER		
FT	SEQUENCE 433 AA; 45935 MW; 90507B8897D943C0 CRC64;		
SEQ			
Query Match 7.6%; Score 132; DB 1; Length 433;			
Best Local Similarity 19.1%; Pred. No. 0.58;			
Matches 48; Conservative 52; Mismatches 107; Indels 44; Gaps 6;			
Qy	115 TAIRDGTGPSQVVRHMQQQIQOQFGEENTR-LHTAVENLKVAVNELSEQINOLKQLHTRLSL 173		
Db	172 TAQQVASSAQVADTQSRAKVGEDGREAQAIAEMSAIEATGETVEEINALDELDE 231		
Qy	174 FGD-----RLEANTGDFALIAID-----FQLSLEFFKSVGKVTMLSPFPEKL 216		
Db	232 IGEIVGIVTSIVTEQTNMALNASIEAAHADGDGEGFAVVADEIKGLAEETKKAADIEGR 291		
Qy	217 AQSLEKTFESQAVQAMSSVTETLNALKELITENKTVIEQLKADAKAOLREEQVRFLEK 276		
Db	292 IEAIEQEQ-AGDTVTMETSTRTIGETGVSTVEETVDALETIVE-----YTEVDVTGIE 343		
Qy	277 RKQLEEEACSTLSHSIATLQESTTLLKDDSTNLHAVESRLIGVMVDGAEASSTVEASOD 336		
Db	344 IDRATEEQARTAQDVMGTIDDLTISQOTAT-----EADTVAGAAOD 385		
Qy	337 DSAQPDENQS 347		
Db	386 QASASIEEVS 396		
RESULT 15			
MYS3_HYDAT			
ID	MYS3_HYDAT	STANDARD;	PRT; 539 AA.

[illegible]

Search completed: August 13, 2002, 09:30:42
Job time: 378 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 09:21:49 ; Search time 23.55 Seconds
(without alignments)
368.199 Million cell updates/sec

Title: US-09-673-763-8
Perfect score: 1733
Sequence: 1 MFTVSDNTSPVISRASPTF.....DDSAQPQDNQSDAGEHKDS 355

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A-COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B-COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A-COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B-COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS-COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	DB ID	Description
1	140	8.1	1057	4	US-09-541-782-10	Sequence 10, Appl
2	131	7.6	2482	1	US-08-328-254-6	Sequence 6, Appl
3	131	7.6	3248	1	US-08-353-700-1	Sequence 1, Appl
4	131	7.6	3248	5	PCT-US95-16216-1	Sequence 1, Appl
5	126	7.3	2101	1	US-08-466-390-4	Sequence 4, Appl
6	126	7.3	2101	1	US-08-470-950-4	Sequence 4, Appl
7	126	7.3	2101	1	US-08-467-781-4	Sequence 4, Appl
8	126	7.3	2101	1	US-08-195-487-4	Sequence 4, Appl
9	126	7.3	2101	2	US-08-483-924-4	Sequence 4, Appl
10	126	7.3	2101	4	US-09-452-294-1	Sequence 4, Appl
11	126	7.3	2101	5	PCT-US93-06160-4	Sequence 1, Appl
12	125.5	7.2	1184	4	US-09-541-782-2	Sequence 4, Appl
13	122.5	7.1	580	2	US-08-591-079-2	Sequence 4, Appl
14	119	6.9	976	4	US-09-104-324B-4	Sequence 2, Appl
15	118	6.8	1388	2	US-08-328-254-7	Sequence 4, Appl
16	117.5	6.8	1388	2	US-08-685-576-1	Sequence 7, Appl
17	117	6.8	1312	2	US-08-592-126-148	Sequence 1, Appl
18	117	6.8	1312	2	US-08-687-080-51	Sequence 148, App
19	116	6.7	576	2	US-08-533-306A-2	Sequence 51, Appl
20	116	6.7	576	2	US-08-742-923A-2	Sequence 2, Appl
21	113.5	6.5	1068	4	US-09-085-199B-11	Sequence 2, Appl
22	113.5	6.5	1354	3	US-08-685-576-1	Sequence 11, Appl
23	112	6.5	546	2	US-09-067-351-1	Sequence 2, Appl
24	112	6.5	546	4	US-09-360-490-1	Sequence 1, Appl
25	112	6.5	1388	2	US-08-685-576-4	Sequence 1, Appl
26	109	6.3	896	1	US-08-095-737-2	Sequence 4, Appl
27	109	6.3	896	1	US-08-480-145-2	Sequence 2, Appl

28 109 6.3 896 2 US-08-477-389-2 Sequence 2, Appl
29 108.5 6.3 1713 3 US-08-600-982-24 Sequence 24, Appl
30 108.5 6.3 1713 5 PCT-US94-10261A-24 Sequence 24, Appl
31 108 6.2 1939 4 US-09-310-187A-1 Sequence 1, Appl
32 107.5 6.2 729 1 US-08-971-937-2 Sequence 2, Appl
33 107.5 6.2 729 2 US-08-812-533-2 Sequence 2, Appl
34 107.5 6.2 770 1 US-08-445-135-2 Sequence 2, Appl
35 107.5 6.2 2285 4 US-09-308-375-2 Sequence 2, Appl
36 107 6.2 414 5 PCT-US93-03077-3 Sequence 3, Appl
37 107 6.2 1093 5 US-08-533-306A-6 Sequence 1, Appl
38 106.5 6.1 816 2 US-08-742-923A-6 Sequence 6, Appl
39 106.5 6.1 816 2 US-08-533-306A-6 Sequence 6, Appl
40 106.5 6.1 885 2 US-08-742-923A-4 Sequence 4, Appl
41 106.5 6.1 1388 4 US-09-572-191-2 Sequence 2, Appl
42 106 6.1 1886 4 US-08-938-105-3 Sequence 3, Appl
43 106 6.1 401 2 US-08-591-079-4 Sequence 4, Appl
44 105 6.1 608 2 US-08-736-770-1 Sequence 1, Appl
45 105 6.1 608 2 US-08-736-770-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-541-782-10
; Sequence 10, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/541.782
; CURRENT FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1057
; TYPE: PRT
; ORGANISM: H.sapiens
US-09-541-782-10

Query Match 8.1% Score 140; DB 4; Length 1057;
Best Local Similarity 21.7%; Pred. No. 0.00047;
Matches 56; Conservative 56; Mismatches 100; Indels 46; Gaps 10;
Qy 113 VITAIKDTGP-----SQVVRHMKQIQQFGEENTRLHTAVENLKAVNVELSEIQNOLKQ 166
Db 298 VITALVERTPHVPYRESKLTIRLQDSLG--GRTRTSI---IATISPASLNLEETLSTLEY 352
Qy 167 LHRLSDFGDRLEANTG-DFTALIADFQLSLEEK-----SVGKVTMTLSPFEKLAQSLK 221
Db 353 AH-RAKNILNKPENQKLTAKLKEYTEIERLKRDLAAAREKNGYVISEENFRVMSGK 411
Qy 222 ETFQEAQVQAMSSVTELRNLNALKELITENKTIVPOLKADAOQLREEQVRFLEKKRQEL 281
Db 412 LTQEEIIVELIEKIGAVEELNRVTELFMDNKNELDOCKSLQNTQLETTOKHLQET 471
Qy 282 EACSTLSHSIATLQES--TTLLKDSITNLHAYESRLIGVMVDGAESSVFEASQD--- 336
Db 472 K-----LQLVKKEYITSALSTEELKHAASALL-----NTVEETTKDVSC 512
Qy 337 -----DSAQPDQENQSDA 349
Db 513 LHSKLDKRRKAVDQHNAAEA 530

RESULT 2
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022

```

Query Match          7.6%; Score 131; DB 1; Length 2482;
Best Local Similarity 21.9%; Pred. No. 0.011;
Matches 64; Conservative 92; Indels 92; Gaps 10;

QY 140 NTRLHTAVENLKANVVELSEIQNLKQLH-----TRLSDGDRLEANTGDTFTALIADFOLS 195
    ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 214 NSDLQKQCEELVQIKGEILEENLMAEQMHQSFAETSQRISKLOEDTSAHQNVVAETLSA 273
    ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

196 LE----EFKSVGTKVETNMLSPFEKLAQS-----LKTFFSOEAVQAMMSSVTE 238
    || ||: ||||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 274 LENKEKELQLLNDKVETQBQAEIQELKLSNHLLEDLSKELQLLSETLSLEKKE--MSSIIS 331
    || ||: ||||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

QY 239 LRTNLNALKELITENTKVIEQLKADAQREQVRFLK-----LLKDS----- 276
    || ||: ||||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 332 L--NKREIEELTQENGTLKE---INASLNQEKMNLIQKSESFANYIDEREKSISELSDQY 386
    || ||: ||||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

QY 277 -----RKQELFEAGSTLSHSTATQEST-----LLKDS----- 306
    || ||: ||||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 387 KOEKLILLQRCGETGNAYEDLSQKYKAAQEKNSKLECLLNCTSCENRKNLEQLOKEAF 446
    || ||: ||||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

QY 307 -----TNLHAVESRLIGVMVDGAGSSTVEEASQDDSAQPDENQSDAG 350
    || ||: ||||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 447 AKEHQEFUTLKFAAEARNQNLML----ELETVQOALRSEMTDQNNKSEAG 494
    || ||: ||||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

RESULT 3
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A

```

```

Query Match          7.68; Score 131; DB 1; Length 3248;
Best local similarity 21.98; Pred. No. 0.016;
Matches 64; Conservative 44; Mismatches 92; Indels 92; Gaps 10;

Qy 140 NTRLHTAVENLKNYVSELSEQINQLKQLH-----TRLSDGFRLEANTGDTALADFOLS 195
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 846 NSDLQKQCEELVQIKGEIEENLMKAEQMHQSFVAETSQRISKLOEDTSAHONYVAETLSA 905
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Qy 196 LE-----EPKSVGTKVETMLSPFEKLAQS-----LKETFSQEAIVQAMMSSVTE 238
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 906 LENKEKEQLQLNDKVTETQAEIQELKKNHLLDLSLKEQLLLSETLSLEKKE--MSSIIS 963
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Qy 239 LRTNLNALKLITENKTVIEQLKADALREBOVREK-----276
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 964 L--NKREITEETQENGTLIKE--INASLNQEKMNLIQKSESFANYIDEREKSISELSDOY 1018
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Qy 277 -----RKQELEACSTLSHSTATQEST-----LLKDS-----306
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1019 KOEKLILLQRCETGNAYEDLSQYKAAQEKNSKLECLLNECTSLCENRKNKEQLKEAF 1078
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Qy 307 -----TNLHVESRLIGVMVQGAESTVEEASODDSAPODENQSDAG 350
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1079 AKEHQEFLLTKLAFAEERNONLML-----ELETVOQLRSEMTDNQNNKSEAG 1126
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 4
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4

```

RESULT 3
US-08-353-700-1
: Sequence 1, Application US/08353700
: Patent No. 5539919
: GENERAL INFORMATION:
: APPLICANT: YEN, TIMOTHY J.
: APPLICANT: RATTNER, JEROME B.
: TITLE OF INVENTION: NUCLEIC ACID ENCODING A

/ GENERAL INFORMATION:
 / APPLICANT: TOUKATLY, GARY
 / APPLICANT: LIDGARD, GRAHAM P
 / TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
 / TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
 / NUMBER OF SEQUENCES: 6
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: TESTA, HURWITZ & THIBEAULT
 / STREET: 125 HIGH STREET
 / CITY: BOSTON
 / STATE: MA
 / COUNTRY: USA
 / ZIP: 02110

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,950
FILING DATE: 06-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-470-950-4

Query Match 7.3%; Score 126; DB 1;
Best Local Similarity 22.5%; Pred. No. 0.024;
Matches 55; Conservative 48; Mismatches 97;

QY 128 HMKQIQOQFGENTRLHTAVENLKVAVNELSEQINQLQLHTRL
Db 456 HFEEKQQLSSLTLDLQSSISNLS----QAKEELEQASQAH----
QY 184 DFTALIADFLSLSEFKSVCTVMTLSPFFKLAQSLKETFSOE
Db 506 EULTLNATIQQQOQELAGLQKAQEKQA---QLAQTLLQ--QEG
QY 244 NALKELITENKTVIEQLKADQLREEQVRFLEKKRQBLEECACST
Db 558 SSLQKQEQLKVAEKQEAATRODHAQQL-----ATAAEEREAS
QY 304 DSTTNLHAVERSL-IGVWDQGAESTVSE-----
Db 612 EKAAKLEILOQLOOVANEARDSAQTSTVQAQKAEKALSRKVEEEL
QY 348 DAGE 351
Db 672 QVAE 675

RESULT 7
US-08-467-781-4
; Sequence 4, Application US/08467781
; Patent No. 5780596
; GENERAL INFORMATION:
; APPLICANT: TOURKATLY, GARY
; APPLICANT: LIGDARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE M
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HORWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.2
; CURRENT APPLICATION DATA:

```

```
Query Match          7.3%; Score 126; DB 2; Length 2101;
Best Local Similarity 22.5%; Pred. No. 0.024;
```

QY	128	HKQOIQOFGENTRLHTAVENLKAVNYVELSEQINQLKQLTRLSDFGDRLEAN----	TG	183
Db	456	HFEEEQOOLSSLITDQSSISNLS-----QAKEELQASQAH-----	CARLTAQVASLTS	505
QY	184	DFITALIADFOLSLFEKFSVGTVETMLSFFPKLQAQSLKETFSQEAQVAMMSVTELRTNL	243	
Db	506	ELTTLNATIQOQDQELAGLKQAQAEKQA---OLAQTLQO---QEQASQGLRHQVEQLS---	557	
QY	244	NALKELITENKTVIEQLKADQAEQVRFLEKRRKQEELEACSTLSHSIAFLQESTLLK	303	
Db	558	SSLKQEQOLKEVAEKQEAATQDHAQOL-----ATAAEEREASLUREDAALKOLEALEK	611	
QY	304	DSTTNLHVESRL-IGVMVQDQASSIVE-----EASODDSAQPDENQS	347	
Db	612	EKAAKLETLQOOLQVANEARDSAQTSVTQAREKAELSRKVEELQACVETARQEQHEAQA	671	
QY	348	DAGE	351	
Db	672	QVAE	675	

```

RESULT 10
US-09-452-294-1
; Sequence 1, Application US/09452294
; Patent No. 6287790
; GENERAL INFORMATION:
; APPLICANT: Lelievre, Sophie
; APPLICANT: Bissell, Mina
; TITLE OF INVENTION: UTILIZATION OF NUCLEAR STRUCTURAL PROTEINS FOR TARGETED
; TITLE OF INVENTION: THERAPY AND DETECTION OF PROLIFERATIVE AND
; TITLE OF INVENTION: DIFFERENTIATION DISORDERS
; FILE REFERENCE: IB-1454- Sequence Submittal
; Patent No. 6287790
; CURRENT APPLICATION NUMBER: US/09/452,294
; CURRENT FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: 60/110,420
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 1
; LENGTH: 2101
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-452-294-1

```

Query Match	7.3%;	Score 126;	DB 4;	Length 2101;
Best Local Similarity	22.5%;	Pred. No. 0.024;		
Matches	55;	Conservative 48;	Mismatches 97;	Indels 44;
Gaps	9;			

QY	128	HKQKQIQQGEENTRLHTAVENLKVNVLSQINQLKQLHTRLSDFGDRLEAN----	TG 183
		: : : : : : : : : : : : : : : : : : : : : :	
Db	456	HFEEXKQQLSSLLTDQSSISNLS-----QAKEELQASQAH-----	CARLTAQVASLTS 505
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
QY	184	DETLAIADPQLSLEEFKSVGTVMLSPFELQAQSLKETFSQEAQVAMMSVVELRNAL	243
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
Db	506	ELTTLNATIQQDQELAGLKQAKQKA---QLAATLQ---QEQASQSLRHQVOLS---	557
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
QY	244	NALKELITENKIVIQLKADALREQVRFLKRRKQEELEACSTLHSHSIATLQESTTLK	303

Db 558 SLLKQEQQLKEVAEQEATRODHAQOL-----ATAAEEREASLRERDAALKQLEALEK 611
QY 304 DSTTNLHAVESRL-IGVMVDGAGESTVE-----EASODDSAQPODENOS 347
Db 612 EKAACLEILQOOLQVANEARDSAQTSTVQAQREKAEKSRKVBELQACVETARQEQHEAQA 671
QY 348 DAGE 351
Db 672 QVAE 675
RESULT 11
PCT-US93-06160-4
; Sequence 4, Application PC/TUS9306160
; GENERAL INFORMATION:
; APPLICANT: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06160
; FILING DATE: 19930621
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ. EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US93-06160-4

Query Match 7.3%; Score 126; DB 5; Length 2101;
Best Local Similarity 22.5%; Pred No. 0.024;
Matches 55; Conservative 48; Mismatches 97; Indels 44; Gaps 9;
QY 128 HMKQIQQFGGEENTRLHTAVENLKAVNVLSLSEQINOLKOLHTRLSDFGDRLEAN-----TG 183
Db 456 HFEEKQQLSSLLTDLQSSISNLS-----QAKBELEQASQA-----GARLTAQVASLTS 505
QY 184 DFTALADFOLSLEEFKSVGTVMLESPFEKLAQSLKFTFQEAQVQAMMSSVTELRTNL 243
Db 506 ELTTLNATIQOQDELQAGLQKQAKQQA-----QLAQTLLQ--QEQAQSGLRHQVEQLS--- 557
QY 244 NALKELITENKTVIEOLKADALREQRVFLKRRKQLEEEACSTLSHSTATIQESTTLK 303
Db 558 SLLKQEQQLKEVAEQEATRODHAQOL-----ATAAEEREASLRERDAALKQLEALEK 611
QY 304 DSTTNLHAVESRL-IGVMVDGAGESTVE-----EASODDSAQPODENOS 347
Db 612 EKAACLEILQOOLQVANEARDSAQTSTVQAQREKAEKSRKVBELQACVETARQEQHEAQA 671
QY 348 DAGE 351

Db 672 QVAE 675
RESULT 12
US-09-541-782-2
; Sequence 2, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/541.782
; CURRENT FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: A. nidulans
; US-09-541-782-2
Query Match 7.2%; Score 125.5; DB 4; Length 1184;
Best Local Similarity 19.8%; Pred No. 0.011;
Matches 49; Conservative 59; Mismatches 111; Indels 29; Gaps 7;
QY 119 DGTFSQVVRHMKQIQQFGGEENTRLHTAVENL-----KAVNVLSLSEQINOLKOLHTR 170
Db 697 NGLSAAARLSEEVIGFTQLHSOLTSFNNLGKDLKSIFETMATHLSQKNEINLRAE 756
QY 171 LSDFGDRLEANTGDTTALADFOLSLEEFKSVGTVMLESPFEKLAQSLKFTFQEAQV 230
Db 757 LQSSNRQNIETTHKSAHLA--QAIEEHVAAEAREILMSQIKALVEESR---QKQFA 810
QY 231 AMMSSVTELTNLNALKELITENKTVIEOLKADALRE---EQVFLKRRKQLEEEACST 287
Db 811 RLRAKIDGVVTEISASGDMLEQATT-----QHDROIDEMWFKSEQFAKDVNASKDEIRTK 865
QY 288 LSHSIATLQESTTLKDDSTNLHAVESRLIGVMVDGAGESTVFEASODDSAQPODENOS 347
Db 866 LQNDWEAFDQORNSTIRKATESVHKETVRIYDVQVDDMGROM---EALDDFVARQSN-- 920
QY 348 DAGEKDS 355
Db 921 --GRYRDA 926

RESULT 13
US-08-591-079-2
; Sequence 2, Application US/08591079
; Patent No. 5972899
; GENERAL INFORMATION:
; APPLICANT: Zychlinksky, Arturo
; APPLICANT: Chen, Yajing
; TITLE OF INVENTION: Apoptosis Induced by Shigella Ipab
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/591.079

;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Livnat, Shmuel
;; REGISTRATION NUMBER: 33,949
;; REFERENCE/DOCKET NUMBER: 15661-20017.00
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 887-1500
;; TELEFAX: (202) 887-0764
;; TELEX: 90-4030 MRSNFOERSHSH
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 580 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-591-079-2

Query Match
Best Local Similarity 7.1%; Score 122.5; DB 2; Length 580;
Matches 54; Conservative 51; Mismatches 98; Indels 35; Gaps 9;
QY 123 SOWRHMKQIQOFGSEENTRLHTAVENKAVNVELSEQINQLKQLHTRLSDFGDRLEANT 182
DB 77 SUTLLIGNLIQILGKSL---TALTN-KITAWKSQQQAQQKNL-----EFSKINTLL 127
QY 183 GDTALIAADPQLSLFEKFSVGTVMKLSPEKLAQSL-----KETFSQAVQAMM 233
DB 128 SETGLTRDYEKQINKLNADSKIKDLENKINQITRLSNLDPSPKKLSREIQ--- 184
QY 234 SSVTELRNLNALKELITENTVIEQ---LKADAQLREQVRFLEKRRKQELFEACSTLS-H 290
DB 185 -----LTIKDAAVKORTLIEQKTLSTHSKLTOKSMQ-LEKEIDSFSAFNTASAE 234
QY 291 SIATLQESTLLKSDTNLHVESRLGVMVQDGAESTVEEASQDSDSAPQDENQSD 348
DB 235 QLSQKSLTGLA-SVTOLMATFQLVKGKNEESLKNDLALFOSLQESRKTEMERKSD 291

RESULT 14

US-09-104-324B-4
; Sequence 4, Application US/09104324B
; Patent No. 6232460
; GENERAL INFORMATION:

; APPLICANT: T. Recl, Ozlem, Sahin, Ugur; Pfreundschuh, Michael
; TITLE OF INVENTION: Methods For Diagnosis And Treating Cancers,
; TITLE OF INVENTION: And Methods For Identifying Pathogenic Markers In A Sample Of
; TITLE OF INVENTION: No. 6232460mal Cells
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fulbright & Jaworski LLP
; STREET: 666 Fifth Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10103

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
; COMPUTER: IBM

; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/104.324B

; FILING DATE: 25-June-1998
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/892,702

; FILING DATE: 15-July-1997
; ATTORNEY/AGENT INFORMATION:

; NAME: Hanson, No. 6232460man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5491
; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (212) 318-3000
;; TELEFAX: (212) 752-5958
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 976 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
US-09-104-324B-4

Query Match
Best Local Similarity 6.9%; Score 119; DB 4; Length 976;
Matches 60; Conservative 37; Mismatches 80; Indels 78; Gaps 10;

QY 139 ENTRLHTAVENKAVNVELSEQINQLKQLHTRLSDFGDRLEANTGDTALIAADPQLSLEE 198
DB 394 EQRLKNEEDQLKILTWELOKKSSELEEM-----TKLTNNKEVELEE 435
QY 199 FKSIVGTVMKLSPEKLAQSLKETFQEAQV---QAMMSSVTELRNLNAL----- 246
DB 436 LKKVLGKEKTLTYENKQFEXIAELKGT-EQELIGLLQAREKEVHDLIEIOLTAITTSQY 494
QY 247 -----KELITE-----NKTVE-----OLKAD--AQLREQVRFLEKRRKQ 279
DB 495 YSKEVKDKLTKTELENEKLNKNTLTSHCNKLSLENKELTQETSDMTLELKNQOEDINNKKQ 554
QY 280 ELEEEACSTLSHSTATLQESTLLKDS-----TTNLHVESRLGVMVQDQ 324
DB 555 E-----ERMLKQLENQETQETQLRNELEYVREELKQKRDVEKCKLDKSENCNHLRKQVE 609
QY 325 AESTVEEASQDSDS 339
DB 610 NKNKYIEELQEQENKA 624

RESULT 15

US-08-328-254-7
; Sequence 7, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:

; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein

; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254

; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239

; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:

; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949

; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 180 amino acids

Tue Aug 13 14:23:42 2002

```

; TYPE: amino acid
; TOPOLOGY: linear
US-08-328-254-7

```

Query Match	6.8%	Score 118;	DB 1;	Length 180;
Best Local Similarity	24.0%;	Prid. No. 0.0033;		
Matches	41;	Conservative	41;	Mismatches 47;
		Indels	42;	Gaps 7;
QY	128	HKQOIQIQGENTPLHTAVENKAVNVELSEGINOLKQLHTRLSDFGDRLEANTGDFTA	187	
		: : : : : : : :		
DB	26	HIAELKRERENOSLKDKNLER-ELQENSEQEL-	61	
		: : : : : : : :		
QY	188	LIADFQLSLEEFKSVGTVKVTMLSPFKLAQSUKETFSQEAV-----QAMMSSVTELRT	241	
		: : : : : : : :		
DB	62	IVLDAENSKAE-----VETLKTQIEAMSLK-VFELDVLTRSEKENLTKQIOEKQ	113	
		: : : :		
QY	242	NLNAUKELITENKTVIEQL-KADQALRREQ---VRFLKREKQOEACSTL	288	
		: : : : : : : : : :		
	114	QSELDKLLSSLEKEQAEQIQKEESTAVEMQLNQKLKELNEAVAL	164	

Search completed: August 13, 2002, 09:24:18
Job time: 149 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 09:24:04 ; Search time 48.57 Seconds
(without alignments)
1264.427 Million cell updates/sec

Title: US-09-673-763-8

Perfect score: 1733

Sequence: 1 MTVSTNDTSPVISRASSPTF.....DSSAQPDENQSDAGEHKOS 355

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1733	100.0	355	2 Q46210	Q46210 chlamydia t
2	448	25.9	225	2 Q32263	Q32263 chlamydia p
3	200	11.5	390	16 Q92828	Q92828 chlamydia p
4	185	10.7	276	16 Q9PKR8	Q9PKR8 chlamydia m
5	166.5	9.6	273	2 Q9RFK7	Q9RFK7 chlamydia t
6	164.5	9.5	273	2 Q9AMA9	Q9AMA9 chlamydia t
7	164.5	9.5	273	2 Q9AM94	Q9AM94 chlamydia t
8	164.5	9.5	273	16 Q84121	Q84121 chlamydia t
9	163.5	9.4	273	2 Q99Q56	Q99Q56 chlamydia t
10	162.5	9.4	273	2 Q9F7K9	Q9F7K9 chlamydia t
11	162.5	9.4	273	2 Q69196	Q69196 chlamydia t
12	162.5	9.4	273	2 Q9AMA7	Q9AMA7 chlamydia t
13	161.5	9.3	273	2 Q9AMB2	Q9AMB2 chlamydia t
14	161.5	9.3	273	2 Q9AMA8	Q9AMA8 chlamydia t
15	161.5	9.3	273	2 Q9AMA6	Q9AMA6 chlamydia t
16	160.5	9.3	273	2 Q9AMB1	Q9AMB1 chlamydia t

17	160.5	9.3	273	2 Q9AMA4	Q9AMA4 chlamydia t
18	157.5	9.1	273	2 Q9AMB0	Q9AMB0 chlamydia t
19	157.5	9.1	273	2 Q9AMA5	Q9AMA5 chlamydia t
20	150.5	8.7	1627	5 Q96200	Q96200 chlamydia t
21	149	8.6	256	2 Q9F7L1	Q9F7L1 chlamydia t
22	149	8.6	533	17 Q9VCP2	Q9VCP2 aeropyrum p
23	149	8.6	612	4 Q96JV2	Q96JV2 homo sapien
24	148.5	8.6	568	10 Q93ZJ6	Q93ZJ6 arabidopsis
25	146.5	8.5	924	5 Q15738	Q15738 dictyosteli
26	145.5	8.4	1732	11 Q54874	Q54874 rattus norv
27	145	8.4	409	16 Q9CKU8	Q9CKU8 pasteurella
28	145	8.4	1690	5 Q4929	Q4929 drosophila
29	145	8.4	1690	5 Q9VJE5	Q9VJE5 drosophila
30	143.5	8.3	1618	11 Q9QX19	Q9QX19 rattus norv
31	143	8.3	1410	4 Q14221	Q14221 homo sapien
32	143	8.3	1411	4 Q15075	Q15075 homo sapien
33	143	8.3	1935	5 Q4934	Q4934 loligo peal
34	141	8.1	678	2 Q9KK49	Q9KK49 streptococ
35	140.5	8.1	403	16 Q9X252	Q9X252 thermotoga
36	140.5	8.1	593	10 Q9AS75	Q9AS75 oryza sativ
37	140	8.1	446	4 Q9RCH4	Q9RCH4 homo sapien
38	140	8.1	770	16 Q9KKP6	Q9KKP6 vibrio chol
39	139.5	8.0	739	17 Q9UYF8	Q9UYF8 pyrococcus
40	139.5	8.0	746	5 Q25561	Q25561 naegleria f
41	138.5	8.0	1641	11 Q88528	Q88528 mus musculu
42	138.5	8.0	1762	10 Q94DC2	Q94DC2 oryza sativ
43	138.5	8.0	1957	5 Q04010	Q04010 onchocerca
44	138.5	8.0	2055	11 Q88938	Q88938 mus musculu
45	137	7.9	592	2 Q00720	Q00720 group g str

ALIGNMENTS

RESULT 1
ID Q46210 PRELIMINARY; PRT: 355 AA.
AC Q46210;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE INCLUSION MEMBRANE LOCALISED PROTEIN.
GN INCA.
OS Chlamydia caviae.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83557;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GPIC;
RX MEDLINE=95302975; PubMed=7783634;
RA Rockey D.D., Heinzen R.A., Hackstadt T.;
RT "Cloning and characterization of a Chlamydia psittaci gene coding for
a protein localized in the inclusion membrane of infected cells.",
RL Mol. Microbiol. 15:617-626(1995).
DR EMBL; L35036; AAC1443.1;
SQ SEQUENCE 355 AA; 38802 MW; 335AAD114D226351 CRC64;

Query Match 100.0%; Score 1733; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 4.3e-93;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MTVSTNDTSPVISRASSPTFGDGHGKDFDNNKIIPISIEAPTSSAAAVGAKTATEPEGRSP	60
Db	1	MTVSTNDTSPVISRASSPTFGDGHGKDFDNNKIIPISIEAPTSSAAAVGAKTATEPEGRSP	60
Qy	61	LLQRICVLVKIIIAIALFVVGIAALVCLYLGVSIVSTPSLILMLAIVLVSFVIVITAIRDG	120
Db	61	LLQRICVLVKIIIAIALFVVGIAALVCLYLGVSIVSTPSLILMLAIVLVSFVIVITAIRDG	120
Qy	121	TPSQVVRHMKQIQIQFGEENTRLHTAVENLKVNVVELSEQINQLKQLHTRLSDFGDRLEA	180
Db	121	TPSQVVRHMKQIQIQFGEENTRLHTAVENLKVNVVELSEQINQLKQLHTRLSDFGDRLEA	180

Tue Aug 13 14:23:44 2002

QY 181 NTGDTALADFQSLSEEFKSVGKVTMLSPFKLAQSLKETFQEAQVAMSSVTELR 240
 DB 181 NTGDTALADFQSLSEEFKSVGKVTMLSPFKLAQSLKETFQEAQVAMSSVTELR 240
 QY 241 TNLNALKELITENKTVEIQKADQAUREQVRFLEKRRKQELBEACSTLSHSIATLOESTT 300
 DB 241 TNLNALKELITENKTVEIQKADQAUREQVRFLEKRRKQELBEACSTLSHSIATLOESTT 300
 QY 301 LKDDSTNLHVESRLIGVMVQDGAESSTVEEASQDSDAQPDENOSDAGEHKDS 355
 DB 301 LKDDSTNLHVESRLIGVMVQDGAESSTVEEASQDSDAQPDENOSDAGEHKDS 355
 RESULT 2
 Q53263 Q53263 PRELIMINARY; PRT; 225 AA.
 AC Q53263;
 01-NOV-1996 (TReMBLrel. 01, Created)
 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
 DE HYPOTHETICAL 25.2 KDA PROTEIN (FRAGMENT).
 OS Chlamydia psittaci (Chlamydia phila psittaci).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia phila.
 OX NCBI_TaxID=83554;
 RN [1]
 RX MEDLINE=92333842; PubMed=1821438;
 RA Sato C., Katumata A., Takashima I., Hashimoto N.;
 RT "Nucleotide sequence of a gene encoding a new genus specific protein
 of Chlamydia psittaci".
 RL Jpn. J. Vet. Res. 39:159-165(1991).
 DR EMBL; S40167; AAB22558.2; -;
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 225 AA; 25165 MW; 74C3D808E749A8FA CRC64;

Query Match 25.9%; Score 448; DB 2; Length 225;
 Best Local Similarity 46.9%; Pred. No. 7.le-19;
 Matches 92; Conservative 44; Mismatches 50; Indels 10; Gaps 2;
 QY 158 SEQINOLKOLHRLSDGDRLEANTGDTALADFQSLSEEFKSVGKVTMLSPFKLA 217
 DB 1 ANQISELTQLHEKLSDFGNKLETCGFDLISEFKVNLEAFKSVGKVTETIISPFERLA 60
 218 QSLKETFQEAQVAMSSVTELRNLNALKELITENKTVEIQKADQAUREQVRFLEKR 277
 61 ESLTGVSFSDAVKNMVDAYSALRVEMETLKQHVESRAVLQQLQSDAQLREQHLLYLEQR 120
 278 KQLEEEACSTLSHSIATLOESTTLLKDDSTNLHVESRLIGVMVQDGAESSTVEEAS 334
 121 KQLEEAVCSTLSASIEQ-----LRSSSTNLQAVESRIVSAVSEDTRRVSLTSTETAD 173
 335 QQDSDAQPDENOSDAG 350
 DB 174 QGDLRDPGDRYGWG 189

RESULT 3
 Q92828 Q92828 PRELIMINARY; PRT; 390 AA.
 AC Q92828;
 01-MAY-1999 (TReMBLrel. 10, Created)
 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE SIMILARITY TO CT119 INCA (CPJ0186 PROTEIN).
 GN CPN0186 OR CPJ0186 OR CP0581
 OS Chlamydia pneumoniae (Chlamydia phila pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia phila.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.

PC STRAIN=CWL029;
 RX MEDLINE=99206606; PubMed=10192388;
 RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis".
 RL Nat. Genet. 21:385-389(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heideberg J.F.,
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
 pneumoniae AR39".
 RL Nucleic Acids Res. 28:1397-1406(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 from Japan and CWL029 from USA".
 RL Nucleic Acids Res. 28:2311-2314(2000).
 DR EMBL; AE001605; AAD18339.1; -;
 DR EMBL; AE002217; AAF38399.1; -;
 DR EMBL; AP002545; BAA98396.1; -;
 DR HSSP; P05412; 1FOS.
 DR TIGR; CP0581; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 390 AA; 42507 MW; F78B8469760A4FD4 CRC64;

Query Match 11.5%; Score 200; DB 16; Length 390;
 Best Local Similarity 23.0%; Pred. No. 0.00031;
 Matches 87; Conservative 64; Mismatches 142; Indels 86; Gaps 13;
 QY 8 TSPVISRASPTGDKGKDFDNKKIIPISTEAPTSSAAAVGAKTAIEPEGRSPLLQRIY 67
 DB 2 SSPVNTTSPAPN-----IPAPTTP-----GIPTT---KPRSFIEKVI 38
 68 LVKTIATAIAFLVVGIAALVCLYLGSVIS-----TPSL-ILMLAIVSFVIVITADTGT 121
 39 ---IVAKYILFAI---AATSGALGTILGSLGALTGGIGIALLVFFVSMVLGLILKDSI 92
 122 PSQVVRHMKQIQOFGNE-----TLHTAVENLKAVN----- 154
 93 SGGEERLREVSFTSENQRLTVITTTLEFEVKDLKAADQLTLEIEAFRNGENLKT 152
 155 -VELSEQINOLKOLHRLSDGDRLEANTGDTALADFQSLSEEFKSVGKVTMLSP 212
 153 AEDLEEQVSKLSEQLALERINLIQANAGDAQEISSELKKLISGWDKSVVEQINTSI-- 210
 213 FEKLAQSLKETFSQEAQVAMSSVTELRNLNALKELI-----TENKTVEIQKADAQ 265
 211 -----QALKVLILGQEWQVQEAQTHVKAMQEQIQAQAEILGHMNSTALOKSVENLLVQDQ 265
 266 LREEQVRFLEKRRKQELBEACSTLSHSIATLOESTTLLKDDSTNLHVESRLIGVMVQDGA 325
 266 ALTRVVGELLENKLSQACSAALROEIEKLAQHETSLQORIDAMLAQBNL-----A 317
 326 ESSTVEASQDSDAQPD 344
 318 EQVTALEKMKQEAQKAASE 336
 RESULT 4
 Q9PKR8 PRELIMINARY; PRT; 276 AA.
 ID Q9PKR8
 AC Q9PKR8;

